

Open Research Online

The Open University's repository of research publications and other research outputs

Metascientific aspects of topoi of spaces

Thesis

How to cite:

Melville, Charles Edwin Brian (1991). Metascientific aspects of topoi of spaces. MPhil thesis The Open University.

For guidance on citations see [FAQs](#).

© 1990 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000fc68>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

UNRESTRICTED

METASCIENTIFIC ASPECTS OF TOPOI OF SPACES

Charles Edwin Brian Melville

A thesis submitted to the Department of Design at
The Open University
for the degree of Master of Philosophy.

August 1990

Date of submission : 7th August 1990
Date of award : 21st February 1991

ProQuest Number: 27758405

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent on the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 27758405

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All Rights Reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ABSTRACT

METASCIENTIFIC ASPECTS OF TOPOI OF SPACES

C.E.B.Melville

August 1990

This thesis presents a study of the importance of topoi for Science. It is argued that whenever the concept of space enters the practice of Science then formal (mathematical) theories should be interpreted in a topos of spaces. It is claimed that these topoi encode knowledge of space arising directly out of the needs of Science, in that the constitutive questions of the Sciences can be traced back to their leading knowledge interests and these determine the role of mathematics as a methodical device. In the Natural Sciences the constitutive questions involve the study of non-intentional objects, in terms of a causal nexus to be explained geometrically, and this facilitates the introduction of geometric objects as the methodical device for posing questions to Nature. Although the study of intentional subjects in the Human Sciences requires ordinary language, not mathematics, to pose questions to each other, secondary methodological objectifications permit a conception of geometric objects analogous to that of the Natural Sciences. Lawvere's axioms for the gros and petit topoi illustrate attempts to formalise the idea of topoi of spaces, as a rational reconstruction of categories in which geometric objects satisfying the formal theories of Science can be found. The catalysing function of this knowledge of topoi of spaces can lead to a diagnosis of mathematical difficulties caused by a failure to align mathematical conceptions with these topoi. This is illustrated through Varela's use of self-reference in Biology and Atkin's use of algebraic topology in Social Studies.

Acknowledgements

This research was carried out under the auspices of the Centre for Configurational Studies (at the Open University). The Centre is concerned with multi-disciplinary studies of spatial arrangements of objects from designed artifacts to social systems. I am grateful to Professor Cross and his colleagues for providing the opportunity to undertake research about space-in-general for the Sciences. In particular, I am grateful to Dr J.H.Johnson for his constant encouragement and useful suggestions. I am also grateful to Dr C.Earl for many stimulating conversations and a critical reading of various drafts. Finally, I owe a special debt to my wife, Dr Sylvia Melville, for reminding me of the ubiquitous 'rule of metaphor' and providing the essential geometry to support this research.

CONTENTS

	<u>Chapter</u>	page
1	Introduction	5
2	The Effectiveness of Mathematics in the Natural Sciences	17
3	Schematizations of Space, Time, Substance, and Causality	34
4	Quasi-causal and Structural Requirements for Space generated by the Human Sciences	50
5	Some Categorical Concepts	71
6	Type Theory as the Formal	93
7	Key Aspects of Elementary Topoi	112
8	Towards Topoi of Spaces	143
9	Varela and Self-reference in Biological Systems	176
10	Atkin and Simplicial Complexes in Social Studies	190
11	Concluding Comments	214
	Notes	224
	Bibliography	228
	Index of Symbols	234
	<u>List of Figures</u>	
1	Subsystems of Science	10
2	Formal and Conceptual Subsystems	10
3	Apel's Ideal Types	54
4	Picture of Adjunction	80
5	Lawvere's World Picture	93
	<u>List of Tables</u>	
1	Observations of a Falling Stone	29
2	Habermas' Schema	36
3	Examples of Human Science	63
4	Interpretation of Calculus of Indications as Propositional Calculus	179

CHAPTER ONE

INTRODUCTION : METASCIENTIFIC ASPECTS OF

TOPOI OF SPACES

One of the most remarkable developments in modern mathematics during recent decades has been the creation and elaboration of the theory of elementary topoi. If a mathematical category is a 'universe of (mathematical) discourse', then any category which is an elementary topos may be regarded as such a universe "... within which we can carry out constructions in much the same way as we do within the category of sets, and with much the same results" ([JT1] p.47). By well-established abuse, topos theorists usually treat objects in a topos almost as if they were sets. Thus a topos is a set-like category; and just as the familiar category of sets and functions corresponds to classical set theory, so elementary topoi may be said to correspond to intuitionist set theory. Because of the intimate connection between topos theory and set theory (and its logic), there is an obvious danger that topos theory may be dismissed as being of interest only to logicians. Indeed Wraith was to urge that topos theory "... is such a pretty subject that it would be most disappointing if it were not good for anything; all my instincts tell me it will be useful, and not just for applications in logic" ([WR] p.115). If this remark is viewed as prophetic, then it has surely been amply fulfilled in the work of Lawvere and those who have followed him in demonstrating how geometric ideas can be applied through the medium of topos theory in contexts where earlier generations would never have suspected their utility.

Observing that "... some physicists and engineers seem in effect to have the insight that geometrical and physical constructions can be performed, with almost as much freedom as sets can be defined in naive set theory, without ever leaving the realm of smooth objects and smooth maps" ([LWD] p.3), Lawvere has argued that there are some topoi which are not just set-like but also space-like, in the sense that objects in these topoi are actually spaces. He has sought to establish the conceptual basis of these topoi of spaces within a class of topoi known as variable sets. Traditionally, set theory has emphasised the constancy of sets in which all variation is frozen. However, variable sets are construed as sets varying with stages of definition. In the first instance, a topos of spaces may be regarded as a category of variable sets, in which the stages of definition are parameterised by some geometric object (such as a 'point' or 'path'). Furthermore, there are certain properties a topos of spaces often has, and a wise selection of these should serve to distinguish them axiomatically from other classes of topoi. Although Lawvere regards his own attempts to axiomatise what he calls gros topoi of spaces as failing to capture the determinate abstract general relation underpinning such topoi, it is true to say that his axioms list important properties underlying the idea of topoi of spaces. As such they serve a catalysing function in guiding the learning, development, and use of these topoi within mathematics.

The ideas underpinning topoi of spaces have a wider metascientific significance than their presentation as abstract mathematical concepts might lead us to suspect. This should not be too surprising as a primary source for geometry has always been the physics of continuous bodies and fields. I argue that the axioms of

topoi of spaces are not simply axioms of an arbitrary (mathematical) character, a mathematical game as it were; rather they distill the essence of our historical and scientific experience of matter moving in space. The extra-mathematical character of the notions of topoi of spaces can best be summarised in the following dictum : whenever the concept of space enters the practice of Science in a fundamental way, then an appropriate 'universe of discourse' for interpreting formalised fragments of the theories of Science should be a topos of spaces rather than an arbitrary topos or other category. If axioms formalising topoi of spaces encode our scientific experience, then the salient extra-mathematical features of these axioms can best be exhibited by locating their normative function in advising on the character of mathematics required for Science. I claim that the axioms of topoi of spaces can contribute to the growth of knowledge, in that they can serve a catalysing function (summarised in the dictum) by guiding the learning, development, and application of mathematics in the Sciences (see Note 1, p.224). Thus the object of this thesis is to elucidate the metascientific aspects of our knowledge of topoi of spaces by showing that

- 1) knowledge of topoi of spaces encodes knowledge of space which arises directly out of the needs of Science, and
- 2) the catalysing function of this knowledge can lead to rapid diagnosis of certain mathematical difficulties in scientific research practice.

My use of the term 'space' will be very general. On the one hand, I shall think of it as one of the highest universals under which the experience of objects may be grasped. On the other hand,

I shall think of 'space' as just as exactly what it says it is: a continuous extension viewed with or without reference to the existence of objects within it (see Oxford English Dictionary). Thus the animating idea of 'space' is that it has, at least, 'points' and 'paths' to reference the location and movements of the objects of experience. However, my use of the term 'space' is not to be identified with any particular mathematical representation of the term. Rather my concern is with the categories in which such representations may be found. Similarly my use of the term 'science' is also very general. For me, the Sciences are concerned with an adequate understanding (episteme) of the object of enquiry, whether that quest is concerned with 'the' external world (= Natural Sciences) or 'our' social-life world (= Human Sciences). Thus the Sciences are concerned with a rational organisation of knowledge and are instituted as a practice which alters its form in the light of new knowledge.

In his answer to the question 'What is Metascience?', Radnitzky argues that there are many ways of looking at Science as research about research ([RA] pp.x-xiv). He describes a sequence of widening perspectives leading from studies of the logical, semantical, information-theoretical, and epistemological aspects of Science to studies of what Science means for man, all of which might be described as 'Metascience'. The underlying disciplines of such studies range from Formal Logic through Sociology and History to the Philosophy of Science. He argues that although these perspectives complement and mediate each other, there is always a danger of totalising any of them, and he suggests that metascientific knowledge is best developed in a process of tacking between these perspectives. Accordingly he proposes a case for a

new type of study : Metascience as systems-theoretically oriented research into Science conceived with practical intent. In a sense, this conception of Metascience is only a means (in modern dress) of tackling the ancient problem of relating theoretical knowledge (episteme) to practical intelligence (phronesis).

Radnitzky views Science as a knowledge producing and distributing enterprise, and describes it as a system. Component parts of this system are : the producers (researchers), the process of research, the products, the supply of intellectual resources, and Metascience as a feedback mechanism which recycles knowledge about research strategies for improved steering of the research enterprise ([RA] I, pp.1-2). The research enterprise is considered as an adaptive innovative system - "the most innovative system there is" ([RA] p.xiv). For such a system, the key metascientific question is 'How can the research process be facilitated ?'. Thus, one is interested in how research is done, and how it is evaluated. For a Metascience conceived with practical intent, another key question is 'How can the research strategy be improved ?'. Here, the focus is on knowledge, which can be evaluated in terms of its relevance for improved steering of the scientific enterprise. Thus, the normative function of Metascience lies largely in catalysing and advising scientific research. "Ideally, it would supply maps ... or theoretical models to those engaged in the study of a given science or some aspects of the knowledge-producing industries" ([RA] p.xii). In a striking (but perhaps trite) metaphor, Radnitzky likens the metascientist to a management consultant concerned to improve the scientific enterprise. From this perspective Metascience is feedback-controlled knowledge for the improved

steering of Science.

The Metascientific character of this thesis is reflected in its objectives. The dictum, that whenever notions of space enter the theories of Science then interpretation should be in a topos of spaces, is knowledge designed to improve the performance of the scientific enterprise. Essentially this thesis studies a subsystem of Science as indicated in Figure 1.

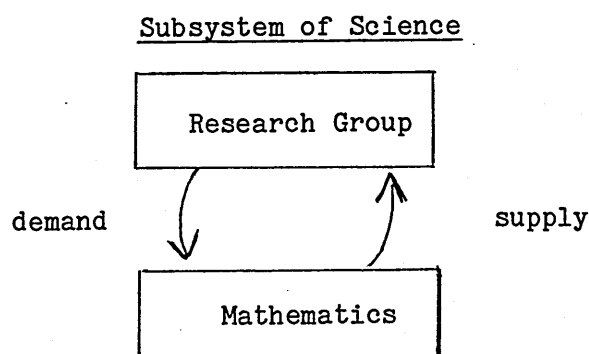


Figure 1

Here, a research group of scientists demand intellectual resources from Mathematics as a component of its products, to which Mathematics attempts to supply appropriate models. A more detailed view of the cycle of interaction is summarised in Figure 2.

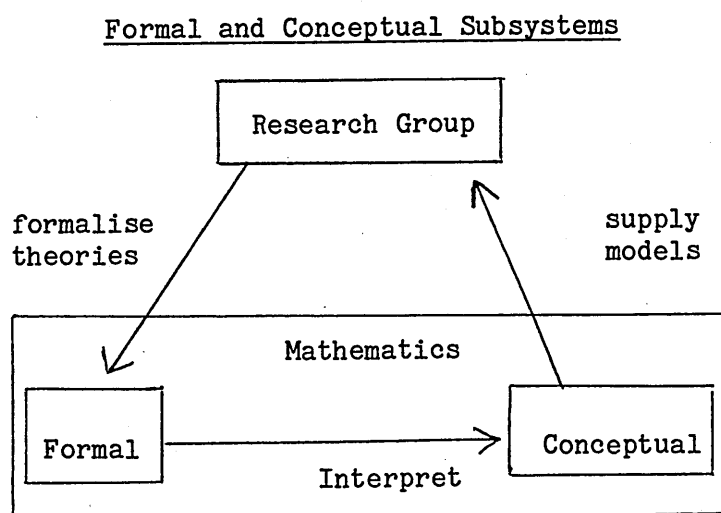


Figure 2

Here, the research group registers its demand for intellectual resources by formalising fragments of its discourse as formal theories expressed in some mathematical/logical language. The Mathematical subsystem is now disaggregated into the Formal and the Conceptual. On the one hand, mathematicians are concerned with the mathematics that can be conceived. On the other hand, their concern with lucidity and precision in expression is reflected in their practice of formalising their conceptions in formal languages. The relation between the Formal and the Conceptual is one in which mathematical concepts are exhibited as models of formal theories in an act of interpretation ([LWA]). Sometimes a variety of models may be exposed, and the supply of these concepts to a research group closes the cycle of interaction. For scientific theories embodying conceptions of space, the knowledge in the dictum is designed to improve the act of interpretation. By delimiting the Conceptual to the class of topoi of spaces, a finer regulation is exerted over the supply of models to a research group. By an understanding of the notions inherent in topoi of spaces, and by feeding this knowledge back into the cycle of interaction indicated in Figure 2 it is possible to improve the performance of the scientific enterprise.

I shall not attempt to make a detailed genetic study of this cycle of interaction. To support my first objective, I shall tack between the Philosophy of Science and Mathematics in a study of more limited scope and warrant. From the perspective of the Philosophy of Science, I use the basic distinctions introduced by Habermas and Apel that the leading knowledge interest in the Natural Sciences is technical, whereas the leading interest in the Human Sciences is practical ([HA1,AP2]). I argue that our

experience in the control and manipulation of moving bodies in space forms the basis of a technical cognitive interest. The schematization of space, time, substance, and causality, where non-intentional objects in which processes are understood in terms of a causal nexus, requires the introduction of a geometric framework where time and space parameterise objectified processes. Thus the Natural Sciences require 'geometric objects' as the essential medium in which to pose questions to Nature. On the other hand, the Human Sciences have a leading knowledge interest in improving communication and understanding between human subjects. Ordinary language, not mathematics, is the essential medium in which to resolve our questions. However, secondary methodological objectifications in the Human Sciences permit the possible introduction of mathematics in two ways. On the one hand, there are circumstances in which quasi-causal Sciences proceed, for one reason or another, as if their research guiding interests were those of the Natural Sciences. On the other hand, there are those Human Sciences which, in the course of their theorising, require precise and formal conceptions of a spatial character expressed in mathematics. This need for (what I shall call for want of a better term) a structural mathematics derives from a sense of structure as the organisation of connected parts. Metaphoric redescription of social cohesion in terms of spatial connectivity underwrites a demand for 'geometric objects'.

From the perspective of Mathematics, I refer to an adjunction between type theory and elementary topoi which enables us to treat the relationship between the Formal and the Conceptual as one of interpreting type theory in an elementary topos ([LS]).

Science needs the formalism of type theory as it is the most powerful representation of the Formal known to us, and is most suitable for the expression of scientific theories. If the theories of Science need to be interpreted in a topos, then the 'geometric objects' required by the Sciences can be found as objects (= spaces) in a topos of spaces. By a 'topos of spaces', I mean a topos in which spaces are objects in which 'geometric elements' (such as 'points' and 'paths') cohere. The crucial moment in my argument, that the ideas underpinning topoi of spaces stem directly from Science's engagement with Nature, lie in the axioms for a gros topos (of spaces), which involve a covariant approach to 'geometric objects' ([LWB]). The observed covariation of space-time events is facilitated by indexing these events with geometric elements in such 'geometric objects'. Furthermore, all attempts to generate other formal systems of topoi of spaces (such as the petit topoi) involve choosing an object (= space) from a gros topos to parameterise the topos. Thus, in some sense, the gros topoi are primary in our attempts to encode knowledge of space arising out of the needs of the Sciences.

To support my second objective, I shall present some limited and partial evidence. My own interests and knowledge were the dominant factor in the selection of the material. I offer some thumbnail sketches of interactions between Research Groups and Mathematics, in which mathematical aspects of the research products have not been cast in terms of topoi of spaces. I argue that attempts to improve mathematical aspects of the research products have been processes of realigning the mathematics in terms of some topos of spaces. I claim, that in the situations I describe, there are already plausible reasons for locating the knowledge of topoi

of spaces in terms of a catalysing function in improving research practice.

Although this thesis is not a work of Mathematics, it is a work about Mathematics. I assume that the reader has served a basic mathematical apprenticeship, and although I have attempted to reduce the mathematical prerequisites to a minimum I assume some knowledge of the elements of category theory. Thus, the ideal reader is either a mathematician interested in the significance of mathematics for Science, or a scientist (with wide mathematical experience) interested in the relevance of mathematics to his research products. The historically-minded reader will be aware that many mathematicians have contributed to the theory of topoi. If I emphasise the work of Lawvere in this thesis, it is not simply due to his leading role in the development of the theory. Lawvere, unlike many other mathematicians, seems always to have theorised with an eye on their wider scientific implications. It is these wider aspects which are highlighted in a work of this nature.

This thesis is divided broadly into three parts, corresponding to the philosophical and mathematical tacking together with empirical studies.

1) Philosophical Tack

Chapter 2 describes the basic distinctions introduced by Habermas and Apel in the theory of leading knowledge interests. I argue that the technical cognitive interest, in which the logic of explanation is reduced to that of prognosis, enables us to understand the introduction of mathematics as a methodical device into the Natural Sciences. The latter is illustrated with the simple example of

Galileo's experiments with falling stones. Chapter 3 establishes that the differences in the constitutive questions can be traced to differing schematizations of space, time, substance, and causality. In particular, I argue that the complete objectification of Nature rests on a metaphorical redescription of natural processes in terms of 'mechanical corpuscular motions'. A technical cognitive interest presupposes that substance and time can be objectified as spaces to explain causality geometrically. Chapter 4 describes a simplified version of Apel's ideal typification of the Human Sciences. This adumbrates two major 'dimensions' of concern :

- i) the distinction between quasi-causal sciences and those purveying good-reason-assays, and
- ii) the contrast between those Human Sciences which evaluate human action in the light of constitutive norms and those which evaluate the norms themselves.

The differing ways in which these ideal typifications can generate a demand for 'geometric objects' is explored, and it is argued that the conceptions of space needed are essentially the same as the Natural Sciences.

2) Mathematical Tack

Chapter 5 outlines the category theory needed to read this thesis. Expert mathematicians need read this only to check any differences with their own use of notation. Of essential importance in much of what follows are the descriptions of various categories of graphs. Chapter 6 describes a simplified version of type theory as our best representation of the Formal. The importance of the internal language of sets and functions is stressed. However, already with Galileo's simple experiments it is argued that Science can benefit from non-classical type theory whenever geometric explanations of motion are required, and taking this seriously rules out the

category of sets as a suitable version of the Conceptual. Chapter 7 describes the set-like categories known as elementary topoi, and shows how, through Kripke-Joyal semantics, theories may be interpreted in variable sets. The topological ideas inherent in Grothendieck topoi are outlined. Chapter 8 is the crucial Chapter in the thesis. It is argued that Lawvere's axioms for a gros topos encode our most general notions of space for describing motion, and that the gros topoi include the familiar examples of spaces known to Science. It is argued that attempts to formalise the notion of topoi of spaces have their starting point in the gros topoi. This is illustrated through the petit topoi.

3) Empirical Studies

The idea that one should be in a topos of spaces whenever notions of space enter the theory can be put to a test. I illustrate the struggle to realign mathematical conceptions in terms of topoi of spaces through two studies :

- i) In Chapter 9, I examine Varela's attempt to import notions of self-reference into Biological Systems ([VG]), and conclude that his revised conceptions are objects in a gros topos.
- ii) In Chapter 10, I examine Atkin's attempts to introduce notions of algebraic topology and cohomology into Social Studies ([AT3]). I note that various attempts to amend his mathematics involve realignment with both the gros and petit topoi.

However, I argue that one cannot understand Atkin's ideas without interpreting the cocycle law in a gros topos of spaces.

Finally, in Chapter 11, I close this thesis with some concluding comments. Notes are at the end of the thesis, together with the Bibliography and an Index of Symbols used in the text.

CHAPTER TWO

THE EFFECTIVENESS OF MATHEMATICS

IN THE NATURAL SCIENCES

It is a common-place that mathematics is useful to us in understanding the world. For example, Wigner has observed that mathematical concepts "... often permit an unexpectedly close and accurate description of the phenomena ..." in the Natural Sciences ([WI2] p.2). In offering us operations (such as collecting, ordering, and measuring) the concepts of mathematics (and particularly geometry) describe entities "... which are directly suggested by the actual world" ([WI2] p.2). Through the symbolic abridgments of mathematics, we may come to understand what the world will let us do to it. "True, in the seventeenth and eighteenth century it was still possible to express and communicate discoveries concerning the 'natural' relations of objects in nonmathematical terms, yet even then - or, rather, particularly then, it was precisely the mathematical form, the mos geometricus, which secured their dependability and trustworthiness" ([KL] p.3). It is no longer possible to conceive of a Natural Science in which the laws of nature must not "... already be formulated in the language of mathematics to be an object for the use of applied mathematics" ([WI2] p.6). Thus, mathematics has come to play a pivotal role in the Natural Sciences.

Why should mathematics be so successful in the Natural Sciences ? Although "... only a fraction of all mathematical concepts are used in physics" ([WI2] p.7), Wigner argues that "...

the enormous usefulness of mathematics in the natural sciences is something bordering on the mysterious and there is no rational explanation for it" ([WI2] p.2). Now good answers to this sensible question ought not to make it an unfathomable mystery. In response to Wigner, Rescher argues that ([RE] pp.122-3) :

"... there is good reason to dismiss all this sort of thing as quite improper and unnecessary mystery-mongering. For man's capacity to discover the laws of nature has a perfectly natural and straightforward evolutionary explanation. There is a good rational basis to support our hope of finding natural laws ... The rationale for this is fundamentally Darwinian: rational guidance is necessary for successful action; successful action is crucial for the survival of creatures ... our survival is indicative of cognitive competence".

To be sure, good reasons can be given from the perspective of Philosophical Anthropology as to the significance of the Natural Sciences for the survival of our species. But there is a sense in which explaining the great utility of mathematics for the Natural Sciences, in terms of species survival, may miss Wigner's point. Wigner was concerned with the effectiveness of mathematics in the Natural Sciences. Although these arguments from Philosophical Anthropology may provide a reasoned account of the significance of the Natural Sciences, they fail to illuminate the effectiveness of mathematics in the nexus between mathematical reasoning and instrumental action. We are left with the impression that mathematics is useful because it is useful.

There is a rational basis for understanding the effectiveness of mathematics in the Natural Sciences in the theory of human cognitive interests (or leading knowledge interests), as developed by Habermas and Apel ([HA1,HA2,AP2,AP3]). Basically this theory postulates three fundamental relations between knowledge and human interests presupposed in any systematic account of the

constitution of meaningful objects of human experience. These are :

- a) the interest in controlling an objectified external environment
(technical cognitive interest),
- b) the interest in communicative understanding (practical or
hermeneutic cognitive interest), and
- c) the interest in critically-emancipatory self-reflection
(emancipatory cognitive interest).

The main advantage of this theory lies in its power in explaining how corroborated knowledge can be released from scientific discourse only for specific pragmatic uses. Thus, although "... all conceivable internal, meaning-constitutive knowledge-interests may be derived ... from the three fundamental knowledge-interests or from possible typical combinations of them ... (and although) ... all three leading internal interests are presupposed in all methodological forms or types of scientific inquiries ...", Apel argues that "... certain types of science (in their constitutive questions and methodical devices) are paradigmatically determined by just one of the three fundamental interests" ([AP2] p.8). Both Habermas and Apel claim that the constitutive questions of the Natural Sciences are paradigmatically determined by a leading knowledge-interest in instrumental action on 'the' external world.

Starting from this viewpoint, it is possible to construct an argument along the following lines.

Firstly, if the relation between knowledge and action (in the Natural Sciences) is limited to that of instrumental action on the external world, then the concepts required (such as space, time, substance, and causality) in the empirical description of that world must already be schematized in such a way as to suit the constitutive questions.

Secondly, this schematization facilitates the introduction of mathematics as a methodical device by which those questions may be posed to Nature. The effectiveness of mathematics is only too reasonable when viewed in this light. It is the device by which we may oblige Nature to answer our questions.

Thirdly, the deployment of mathematics as a methodical device illuminates the demands that Natural Science makes on Mathematics. On the one hand, it explains why only a fraction of all mathematical concepts are required. On the other hand, it explains why Mathematics can only satisfy those demands in terms of the more fundamental ideas of space.

Finally, the concepts that Mathematics can supply can be found as objects in some topos of spaces.

If the above argument is sound, then the first objective of this thesis - topoi of spaces encode notions of space which arise directly out of the needs of Science - can be satisfied, at least for the Natural Sciences. I will develop this argument in Chapter 3. However, as a start on implementing this argument, the remainder of this Chapter will

- a) introduce the theory of leading knowledge interests as a basis for this work, and
- b) show what it means for mathematics to be introduced as a methodical device in terms of a simple example drawn from Galileo's and Newton's observations on falling bodies.

The satisfaction of my first objective vis-à-vis the Human Sciences is somewhat more delicate, and I will defer discussion of this to Chapter 4.

The theory of human leading knowledge-interests is an attempt to unearth the roots of knowledge in the life of the human species ([MC] p.55) :

"It is his (Habermas) central thesis that 'the specific view points from which we apprehend reality', the 'general cognitive strategies' that guide systematic inquiry, have their 'basis in the natural history of the human species'. They are tied to 'imperatives of the socio-cultural form of life'."

A profound consideration of these imperatives results in a distinction between two very basic general cognitive strategies. On the one hand ([MC] p.55) :

"The reproduction of human life is irrevocably bound to the reproduction of the material basis of life. From the most elementary forms of wresting an existence from nature, through ... to the development of ... industry, the material 'exchange process with nature' has transpired in structures of social labour that depends on knowledge that makes a claim to truth. The history of this confrontation with nature ... has the form of a 'learning process'. Habermas's thesis is that the 'general orientation' guiding the sciences of nature is rooted in an 'anthropologically deep-seated interest' in predicting and controlling events in the natural environment, which he calls the technical interest".

On the other hand ([MC] pp.55-6) :

"The reproduction of human life is just as irrevocably based on reliable intersubjectivity in ordinary language communication. The transformation of the helpless newborn into a social individual capable of participating in the life of the community marks his entrance into a network of communicative relations from which he is not released until death. Disturbances to communication in the form of the non-agreement of reciprocal expectations is no less a threat to the reproduction of social life than the failure of purposive-rational action on nature. The development of the historical and cultural sciences from professions in which practical knowledge was organized, transmitted, and applied, brought with it a systematic refinement and extension of the forms of understanding through which intersubjectivity can be maintained. Habermas's thesis is that a general orientation guiding the 'historical-hermeneutic' sciences is rooted in an anthropologically deep-seated interest in securing and expanding possibilities of mutual and self-understanding in the conduct of life. He calls this the practical interest"

Although these arguments are cast in terms of Philosophical Anthropology, they are much better focussed than those of Rescher. The distinction between 'our' social-life world of intersubjectively-valid communication and 'the' external world

based on our experience of moving bodies has its basis in cognition. As Searle puts it ([SE] p.16) :

"... what is the difference between regarding an object as an instance of linguistic communication and not so regarding it ? One crucial difference is this. When I take a noise or a mark on a piece of paper to be an instance of linguistic communication, as a message, one of the things I must assume is that the noise or mark was produced by a being or beings more or less like myself and produced with certain kinds of intentions. If I regard the noise or mark as a natural phenomenon like the wind in the trees or a stain on the paper, I exclude it from the class of linguistic communication ..."

This cognitive distinction between linguistic communication and natural phenomena lies at the heart of this theory, and forms the basis for our understanding of the differences between the Natural and the Human Sciences. The constitutive questions of the Natural Sciences incorporate a technical interest in instrumental intervention in natural phenomena. The constitutive questions of the Human Sciences incorporate a practical interest in improving human understanding.

Habermas and Apel also argue that there is a third derivative mode of inquiry of critical reflection, which they call an emancipatory cognitive interest. The general orientation here is in knowledge which emancipates the human subject from pseudo-natural constraints whose power rests in their non-transparency. Some of their critics argue that this particular interest should be merged with that of the practical. This dispute is somewhat tangential to my concerns so I shall not enter this time-honoured suit here, but simply record my belief that "... interest in this kind of understanding is not simply interest in understanding pure and simple, but the more fundamental form of interest in emancipation from the conditions of coercion, which are experienced as failure and alienation and can be criticized as historically

superfluous" ([WE] pp.50-1).

The argument that the constitutive questions of the Sciences derive from leading knowledge interests can be illustrated by the following prototypical examples, which may be taken as paradigmatic:

- a) Mechanics or Continuum Physics representing the Natural Sciences and illustrating a technical interest, and
- b) Philology representing the Human Sciences and illustrating a practical or hermeneutic interest.

A) Technical Interest

The general orientation of the technical interest lies in controlling the objectified external world. But this cannot mean that, as an interest in knowledge, it should assimilate the methodological structure of Science to that simply of technique. What it does mean is that the very possibility of an experimental and empirical-analytical Science is mediated by the methodical device of posing questions to Nature which can receive answers capable of being transformed into if-then rules. In principle, the latter have the potential to supply instrumental schemes for goal-directed action. The paradigm case of Natural Science is that of Mechanics or Continuum Physics, which attempt to rationally reconstruct our experience of moving and deforming bodies through the prediction of space-time events. Here ([HA1] p.308) :

"... theories comprise hypethetico-deductive connections of propositions, which permit the deduction of law-like hypotheses with empirical content. The latter can be interpreted as statements about the covariance of observable events; given a set of initial conditions, they make predictions possible. Empirical-analytical knowledge is thus possible predictive knowledge. However, the meaning of such predictions, that is their technical exploitability, is established only by the rules according to which we apply theories to reality."

Thus Habermas identifies the link between explanation and control

central to the technical interest and lays the basis for knowledge as the technical exploitation of Nature. He argues that ([HA4] p.137) :

"Out of the very procedure with which the validity of law-like hypotheses is checked against experience, there arises the specific achievement of empirical scientific theories : they permit limited predictions of objective or objectified processes. Since we test a theory by comparing the events predicted with those actually observed, a theory which has been sufficiently tested empirically allows us -- on the basis of its general statements, that is its laws, and, with the aid of limiting conditions which determine a case under consideration -- to subsume this case under the law and to set up a prognosis for the given situation. One usually calls the situation defined by the limiting conditions the cause, and the predicted event the effect. If we use a theory in this way to forecast an event, then it is said we can 'explain' this event. Limited prognosis and causal explanation are different expressions for the same achievement of the theoretical sciences."

In this argument (that an explanation is only adequate if it is, at least potentially, also a prediction) the ultimate aim of Natural Science becomes clear; for the meaning of a prediction is its technical application. Our ability to predict the consequences of an intervention into a natural process is a prerequisite for the successful manipulation of it, or, as Comte put it : "From Science comes Prevision, from Prevision comes Control" (quoted in [FA] p.37). Scientific knowledge makes it possible (for those who possess it) to control the objects of the external world. However, the interest in controlling an objectified external world may be sublimated beyond its paradigm of manipulating things and applied to 'realities' such as 'our' social-life world. Clearly there are dangers in knowledge of 'our' social-life world being reduced to that of the manipulation of human subjects.

B) Practical Interest

These dangers can be avoided if we think of the investigation of 'our' social-life world in terms of a practical cognitive interest. It may be called 'practical' to recall the sense of the

Aristotelian term 'praxis'; for ([HA3] pp.8-9) :

"In interactions ... we encounter objects of the type of speaking and acting subjects; here we experience persons, utterances, and conditions which are in principle structured and to be understood symbolically."

The understanding of 'our' social-life world has its counterpart to the verification of law-like hypotheses in the understanding of meaning. Here, the paradigm example is that of Philology, which attempts to interpret texts (written utterances) through understanding the 'rules' governing the text. An interpretation recognises that conceivable meaning is not necessarily intended meaning, and that the 'real' meaning of actions is not necessarily identical with subjectively intended meaning. It strives for the meaning or a range of possible meanings of a text. Ideally it strives for 'the' meaning of a text in terms of 'rules' with which we may understand it. It strives for the 'consensual reading'. In this sense interpretation passes beyond Philology to the striving for a consensus with other people in human interaction.

The wide scope of this interest may be explained in terms of the (Wittgensteinian) insight that children could not learn the meanings of words if they did not learn to use them at the same time as coming to agreement with other members of their speech-community about the paradigmatic evidences of experience and rules of behaviour. Understanding 'our' social-life world is partly understanding the language in which we speak about our world and partly the understanding of the social constraints which limit our actions. The crucial importance for the Human Sciences to incorporate a practical cognitive interest is summed up by Habermas ([HA1] p.176) :

"In its very structure hermeneutic understanding is designed to guarantee, within cultural traditions, the possible action-

orienting self-understanding of individuals and groups as well as reciprocal understanding between different individuals and groups. It makes possible the form of unconstrained consensus and the type of open intersubjectivity on which communicative action depends. It bans the danger of communication breakdown in both dimensions : the vertical one of one's own individual life history and the collective tradition to which one belongs, and the horizontal one of mediating between the traditions of different individuals, groups and cultures. When these communication flows break off and the intersubjectivity of mutual understanding is either rigidified or falls apart, a condition of survival is disturbed, one that is as elementary as the complementary condition of the success of instrumental action : namely the possibility of unconstrained agreement and non-violent recognition. Because this is the presupposition of practice, we call the knowledge-constitutive interest of the cultural sciences 'practical'."

The danger of thinking of the Human Sciences in terms of the empirical-analytical traditions of the Natural Sciences runs the obvious risk of misunderstandings which may easily lead to those 'communication breakdowns' referred to above.

Some words of warning about possible misunderstandings are necessary when considering the theory of cognitive interests. In the first place, if leading knowledge interests constitute the meaningful objects of experience, then they may not be reduced to or simply equated with individual or institutional motivations which may promote (or hinder) the scientific enterprise. It is not to be assumed that the theory relates to the intentions of any individual or group of scientists. Whereas the latter may be considered as causes of the actual performance of a scientific enterprise, leading knowledge-interests only elucidate reasons for the differences in constitutive questions of the Sciences. Although Apel also argues "... it ... may easily be admitted that, in the long run, they fulfill also a causal or energetic function in the evolution of human knowledge in being the force behind at least part of the external motivation of the scientific enterprise" ([AP2] p.4). This latter remark is of considerable metascientific

interest, for it conceives of the normative aspects of the theory as knowledge contributing to the steering of Science. From a metascientific point of view, leading knowledge interests may be construed as the research guiding interests of the Sciences. From the vantage point of the theory, cognitive interests are not conscious interests. Since they determine the meaning of different forms of knowledge prior to intentions, these intentions cannot simply alter that meaning. A part of that meaning is an idea of a relationship between theory and practice, for the conditions of possibility of each form of knowledge determines its pragmatic use. In the second place, these cognitive interests are not to be thought of as exclusive but complementary. There is always a danger that a contrasting correlation of technical / Natural Science and practical / Human Science may lead to an oversimplified view of the matters under consideration. Instead we must conceive of these interests as leading knowledge interests within their respective spheres, in which the contrasting interest forms a complement to it. The reasons for this are not hard to find ([AP2] p.11) :

"Even physicists cannot explain regularities of nature nomologically without presupposing at the same time another type of knowledge, viz. communicative understanding. This other type of knowledge might be considered prescientific from the point of view of the actual knowledge-interest of the physicist ... This phenomenon (of complementarity) obviously shows the genuine origin and function of hermeneutic understanding and interpretation as it is, in principle, not to be replaced by or reduced to nomological explanation, because it is ... presupposed by all kinds of explanation. On the other hand, the fact that all communicative understanding presupposes some knowledge about the objective matters in question, only illustrates the impossibility of conceiving communicative understanding in terms of an objectifying knowledge ... For as understanding is ... about some objective matter in the world, it cannot be replaced and 'scientized' by just describing and eventually explaining certain objective processes to be observed in our communication partners ..."

From these considerations, the understanding of meaning and the explanation of objectified processes form complementary forms of knowledge. They support each other in terms of possible human

knowledge. They only exclude each other as different cognitive interests in terms of the questions they can ask. They simply cannot be reduced to each other.

The thesis, that the Natural Sciences incorporate a leading knowledge interest concerned with the description, prediction, and control of objectified processes, can be illuminated by the following simple example, which has the advantages of not only being familiar but also illustrating what it means to introduce mathematics as a methodical device. The example, I have in mind, is Galileo's observations on freely falling bodies. Wigner notes that Galileo discovered ([WI2] pp.4-5) :

"... that two rocks, dropped at the same time from the same height, reach the ground at the same time. The laws of nature are concerned with such regularities. Galileo's regularity is a prototype of a large class of regularities. It is a surprising regularity for three reasons. The first reason ... is that it is true on Earth, was always true, and always will be true. This property of the regularity is a recognized invariance property and ... without invariance principles similar to those implied in ... Galileo's observation, physics would not be possible. The second surprising feature is that the regularity ... is independent of so many conditions which could have an effect on it ... The irrelevancy of so many circumstances which could play a role in the phenomenon observed has also been called an invariance. However this invariance is of a different character than the preceding one since it cannot be formulated as a general principle ... The preceding two points, though highly significant from the viewpoint of the philosopher, are not the ones which surprised Galileo most, nor do they contain a specific law of nature. The law of nature is contained in the statement that the length of time which it takes for a heavy object to fall from a given height is independent of the size, material and shape of the body which drops. In the framework of Newton's second 'law', this amounts to a statement that the gravitational force which acts on the falling body is proportional to its mass but independent of the size, material, and shape of the body which falls."

Now the first two points are of obvious interest to the Philosophy of Science. Perhaps Wigner's second surprising feature - the irrelevance of so many matters not connected to the understanding of natural phenomena in terms of causal processes - is, when viewed

in the context that natural laws are statements of natural invariances, no more than an anticipation that Natural Science incorporates a technical cognitive interest. Far from being an 'unformulated general principle', the disconnection of irrelevant matters is what makes Natural Science the study of natural phenomena. The limitation of constitutive questions then enables the introduction of mathematics as a methodical device, and it seems to me that this is the third feature which surprised Galileo so much.

Galileo found that universally quantified statements about the 'behaviour' of falling bodies could be interpreted by means of simple geometrical diagrams. By using simple observational equipment, the motion of falling bodies, as observed by Galileo, can be described by the sort of data in Table 1. A stone is dropped from a height of 400 feet, and the following observations are recorded :

Observations of a Falling Stone

<u>Time</u> <u>(in seconds)</u>	<u>Distance fallen</u> <u>(in feet)</u>
0	0
1	16
2	64
3	144
4	256
5	400

Table 1

Now Greenspan shows ([GR] pp.1-2) that the relation between time

and distance fallen can be described by a mathematical equation $s = 16t^2$, where s denotes the distance travelled (in feet) by a falling stone from the place where it was dropped, and t denotes the elapsed time (in seconds). Thus, a stone dropped from a height of 400 feet will hit the ground in 5 seconds. As a natural invariance, the relation between time and distance can be construed as a universally quantified statement of functional dependence,

$$q:[0,\infty] \longrightarrow \mathbb{R},$$

given by

$$s = 16t^2 \quad (1).$$

But the effectiveness of mathematics does not rest merely in a description of the relation between distance travelled and elapsed time. Greenspan also shows by simple arithmetical manipulation, how we can describe the velocity of any falling stone by the equation

$$v = 32t \quad (2),$$

where v denotes the velocity in terms of distance travelled per second. This indicates that the velocity of a body increases at a constant rate as it falls. We arrive at a regularity, when by further manipulation, we discover that

$$a = 32 \quad (3),$$

where the acceleration in velocity, denoted by a , is a constant 32 feet per second per second. Galileo's discovery of this regularity states a law of nature, in which all bodies falling to the Earth fall with the constant acceleration in velocity as described in (3). Yet these descriptions (useful though they are) are only a part of the story. By a chain of mathematical argument (the details of which need not detain us here), we may arrive at the following :

Proposition 1 For any interval $[a,b]$ and for any function

$f:[a,b] \longrightarrow \mathbb{R}$, there is a unique differentiable function

$g:[a,b] \longrightarrow \mathbb{R}$ with a derivative, denoted by g' , equivalent to f

with $g(a) = 0$.

Proof : ([KK2] Th.13.2).

It was Newton's genius to see that (3) was the derivative of (2), and (2) was the derivative of (1). Starting from the regularity in (3), we can predict velocity by integrating (3) over infinitesimal amounts of elapsed time :

$$\int_0^t a \cdot dt = v(t) - v(0) \quad (4).$$

By integrating velocity (2) over elapsed time

$$\int_0^t v \cdot dt = q(t) - q(0) \quad (5),$$

we can predict the distance fallen by the body. Furthermore, the body's motion is described in terms of causal processes. The application of (gravitational) force causes the body to fall. Newton's famous law, force = mass times acceleration, can, as Greenspan shows ([GR] p.4), be given, with the aid of the above proposition, a more concise mathematical formulation. If the force required to move the body is denoted by W , then the work done by gravity in the time interval $[0, t]$ is defined by

$$W = m \int_0^t a \cdot dt \quad (6),$$

where m denotes the (constant) mass of the body. Routine manipulation ([GR] pp.4-5) of (6) shows that the classical formulae for kinetic energy

$$K = \frac{1}{2} m v^2 \quad (7),$$

and potential energy

$$V = 32 m s \quad (8),$$

are essentially the same in that

$$W = \frac{1}{2} m v^2(t) - \frac{1}{2} m v^2(0) \quad (9)$$

$$\text{and} \quad W = 32ms(t) - 32ms(0) \quad (10).$$

Conservation of energy follows immediately from equating W in (9) and (10). If we use the knowledge that the force is gravity, then by (10) gravity is equated with the mass-potential of the body. However, by (9) that is also the energy required to move the body.

Of course, the above remarks do not exhaust the content of Galileo's and Newton's observations in connection with the laws of freely falling bodies. However, in essence they form the basis of a Science of the dynamics of moving (rigid) bodies. Not only are space-time events to be predicted on the basis of natural laws, but an understanding of the causal effect of forces required to manipulate events is the point of any theory. It is well-known that Newton went on to show how to obtain the motions of bodies from a knowledge of the forces acting on them. He demonstrated the possibility of dealing with gravitational systems in a unified way through the methodical device of the mathematics. Newton wished (quoted in [KI] p.513) :

"... we could derive the rest of the phenomena of Nature by the same kind of reasoning from mechanical principles, for I am induced by many reasons to suspect that they may all depend upon certain forces by which the particles of bodies, by some causes hitherto unknown, are either mutually impelled towards one another, and cohere in regular figures, or are repelled and recede from one another."

Thus, Newton's programme forms the historical basis of our reflections on the Natural Sciences, in which the understanding of 'forces' leading to a particular configuration of events requires mathematical and geometrical knowledge. If we could isolate a few basic laws, akin to the law of universal gravitation, then by application of these basic laws to the specification of the ultimate parts of bodies, all of the phenomena of nature could be derived. To be sure, this requires that natural phenomena be

metaphorically redescribed in terms of our experience of manipulating moving bodies.

Already in this prototypical example of Galileo's and Newton's observations on falling bodies, we can discern all the essential ingredients which bear on the effectiveness of mathematics in the Natural Sciences. Firstly, invariant regularities of nature must be derived as laws of nature. Secondly, the observed covariation of natural phenomena can be described in terms of mathematical relationships. Thirdly, the laws of nature must be postulated as nomological (or law-like) hypotheses to be confirmed in the observed covariation of natural phenomena. Fourthly, mathematical reasoning can be used to investigate the geometrical or mathematical commitments of current theory to produce further hypotheses about predictable events for empirical testing. The effectiveness of mathematics in the Natural Sciences rests partly in its potential to describe laws of nature nomologically and partly, through the logic of mathematical reasoning, the opportunities it yields for potential hypotheses which (if confirmed empirically) could extend the theory. Indeed Wigner remarks that : "It is now natural for us to derive the laws of nature and to test their validity by means of the laws of invariance, rather than to derive the laws of invariance from what we believe to be the laws of nature" ([WI1] p.523). But all this seems to depend on the laws of nature falling under the rubric of a cluster of concepts; namely space (distance travelled), time (interval of time), substance (falling body), and causality (gravitational force), and it is to this I now turn.

CHAPTER THREE

SCHEMATIZATIONS OF SPACE

TIME, SUBSTANCE, AND CAUSALITY

"'Space' and 'time' as the Transcendental Aesthetic lays them out, are, in spite of all assurances to the contrary, concepts, or in Kant's expression representations of a representation. They are not intuitive, but rather the highest universals under which the 'given' may be grasped. The fact ... that a given independent of these concepts is not indeed possible, turns givenness itself into something mediated."

T.W Adorno

In the previous Chapter, I outlined a theory of human cognitive interests as a basis for understanding the introduction of mathematics as a methodical device into the Natural Sciences, and presented some circumstantial evidence as to the relevance of the concepts of space, time, substance, and causality. In this Chapter, I shall take up the first three aspects of the argument outlined in the previous Chapter. My aim is to establish that differences in the constitutive questions of the Sciences can be traced to differing schematizations of the concepts of space, time, substance, and causality. I argue that a technical cognitive interest accords a certain priority to the concept of space, and the conceptualisation of space in terms of geometric figures arises directly out of the (constitutive) questions of this interest.

Successful reference to the objects of the world requires

the deictic expressions which involve notions of time and space, as well as articles and demonstrative pronouns. These form a reference system of possible denotations, as they link the levels of intersubjectivity on which human subjects converse and interact reciprocally with the levels of objects about which the subjects make propositions. Habermas puts it in this way ([HA2] p.173) :

"When we identify objects about which we state something (on the basis of experience we have had), we do so either ostensively or by means of names and characterizations. It is true, predicative determinations are not used predicatively in the context of denotative expressions. But a proper functioning system of reference has to have a certain propositional content. This minimum content of properties which objects as such have is the categorial framework for objectivating experienceable happenings as happenings ... The basic notions of substance, space, time, and causality are the minimum conditions for determining a system of reference for objects of possible experience."

However, although this reference system functions both for the domains of objectified processes and intersubjective communication, it is schematized differently for the two domains in question. The underlying rules for identifying things and events are different but complementary with those for identifying subjects and their utterances. Habermas argues ([MC] p.297) :

"The sense of substance and causality, of space and time, is differentiated according to whether these categories are applied to objects within a world or to the linguistically constituted world of speaking subjects itself. The interpretive schema 'substance' has a different meaning for the identity of objects that can clearly be categorized analytically than it does for speaking and acting subjects whose ego-identity cannot be grasped with clear-cut analytical operations. The interpretive schema 'causality', when applied to observable events leads to the concept of 'cause'; when applied to a nexus of intentional actions, it leads to the concept of 'motive'. Analogously, space and time are schematized differently in regard to the physically measurable properties of objects and events than they are in regard to the intersubjective experience of contexts of symbolically mediated interactions. In the first case the categories serve as a coordinate system for observation controlled by the success of instrumental action; in the latter case they serve as a frame of reference for the subjective experience of social space and historical time."

In the one case we have a reference system for empirical descriptions, and in the other a reference system for narratives.

Habermas' schematizations are summarised in Table 2.

<u>Habermas' Schema</u>		
<u>Schema</u>	<u>Technical Interest</u>	<u>Practical Interest</u>
Substance	object	ego-identity
Causality	cause/ explanation	motive/ understanding
Time	{geometric frame of	historical time
Space	reference}	social space

Table 2

To be sure, the category of causality has a different status than the others, for "... the linguistic representatives of the causal relation do not belong to the class of deictic expressions" ([MC] p.429). Nevertheless, for the purposes of my argument, this is not of great importance. What is important is that these two schematizations can be used to contrast the constitutive questions of both the technical and practical interests. These different schematizations are possible because the cognitive content of deictic expressions is ambiguous. We relate articles and demonstrative pronouns to persons as well as things, and these ambiguities need to be explored.

A) Substance

Now Apel argues ([AP2] p.15) :

"A fundamental precondition of the constitution of nature as an object of experimental explanatory physics in the modern sense is ... defined by the possibility of a complete objectification of nature ... whereas all social sciences, even those that provide the basis for social technology, are only allowed to perform secondary methodological objectifications but should not totally objectify

their human subject-objects, lest they are lost as intentional objects of knowledge." (emphasis added).

Essentially, the Human Sciences can never lose sight of the fact that their fundamental questions are concerned with 'our' social-life world of intersubjective communication. Human subjects are conceptualised as ego-identities with intentions. The situation with the Natural Sciences is radically different ([AP2] p.15) :

"For modern physics it was a precondition of its definitive constitution as science at the beginning of the seventeenth century that all kinds of sympathetic-communicative and teleological understanding of nature were definitely renounced (or dispensed with). Only this renunciation of understanding could release nature as a world of mechanical corpuscular motions, and that is to say, as an object of explanation according to causal laws."

This renunciation of understanding can never completely banish the intentional subject of the Human Sciences. However, it is essential (in the Natural Sciences) if the objectification of Nature is to be achieved. Indeed, Wigner's point, concerning the irrelevance of so many matters not connected to the understanding of Nature, is the dispensation of any form of teleological understanding of Nature. The constitutive questions of the Natural Sciences rest on schematizing substance as non-intentional objects.

However, the complete objectification of Nature would also seem to rest on metaphorically redescribing natural processes as a 'world of mechanical corpuscular motions'. For example, sound as an entity or quality is not normally construed as a 'corpuscular object'. Nevertheless, the scientific study of acoustics involves studying the propagation of sound in terms of varying quantities of pitch, volume, and timbre. Essentially, sound is objectified in terms of motion through a 'sound space' of various states of sound. The reader should try whistling a closed sound curve ([NE] p.15).

Metaphor is, as Ricoeur puts it, "... the rhetorical process by which discourse unleashes the power that certain fictions have to describe reality" ([RI] p.7). Most normal conceptual systems are metaphorically structured, in that most concepts are partially understood in terms of other concepts ([LJ] p.56). Essentially we understand our fictions (or narratives) in terms of other fictions. The unfamiliar is redescribed in terms of the familiar.

Unfortunately it is still ([HE3] p.158) "... necessary to argue that metaphor is more than a decorative literary device and that it has cognitive implications whose nature is ..." to redescribe the domain of the explanandum in terms of an explanans transferred from a secondary system. For example, consider the metaphor ARGUMENT IS WAR ([LJ] p.4). The primary system, ARGUMENT, is explained in terms of notions drawn from the secondary system, WAR. Lakoff and Johnson characterise this redescription as follows ([LJ] p.4) :

"... we don't just talk about arguments in terms of war. We can actually win or lose arguments. We see the person we are arguing with as an opponent. We attack his positions and we defend our own. We gain and lose ground. We plan and use strategies. If we find a position indefensible, we can abandon it and take a new line of attack. Many of the things we do in arguing are partially structured by the concept of war. Though there is no physical battle, there is a verbal battle, and the structure of an argument - attack, defense, counterattack, etc. - reflects this. It is in this sense that the ARGUMENT IS WAR metaphor is one that we live by in this culture; it structures the actions we perform in arguing ... The essence of metaphor is understanding and experiencing one kind of thing in terms of another."

The cognitive implications of this for a deductive theory of explanation are considerable. Now, Hesse has given a detailed account of the explanatory function of metaphor ([HE3]). I cannot go fully into this here. But basically her argument is that if metaphor introduces a new language (or way of looking at things) then redescription may amount to an explanation. According to strict deductivist criteria, it must be possible to deduce the explanandum from the explanans. However, recourse to metaphorical

redescription is a consequence of the difficulty of obtaining such a strict deductive relation between explanandum and explanans. This throws into question our categorisation of rationality along deductivist lines. To counter this, Hesse argues that ([HE3] pp.176-7) "... rationality consists just in the continuous adaptation of our language to our continually expanding world, and metaphor is one of the chief means by which this is accomplished". So for example, on the one hand, the meaning of an objectified process lies precisely in the renunciation of an intentional subject; on the other hand, the metaphor, NATURAL PHENOMENON IS MOVING BODY, redescribes the natural phenomenon (explanandum) in terms of an explanans grounded in the 'world of mechanical corpuscular motions'.

B) Causality

The sense of a complete renunciation of understanding of natural phenomenon in terms of teleological understanding is also illustrated by the schema of 'causation'. Natural Science explains natural invariances not as rules that must be followed by Nature as a subject of teleological action ('the apple falls because it loves the Earth'), but as natural laws determining the 'behaviour' of natural phenomena in terms of causal necessity or probability ('the apple falls because of the application of gravitational force'). Apel argues that methodological consequences flowing from this posture can be formulated as the following theses ([AP2] p.16) :

- a) the Natural Sciences postulate nomological (or law-like) hypotheses as explanations of Nature, and test these by communication-free observations;
- b) observed deviations of events postulated in hypothetical laws

must be considered as falsifications of theory;

- c) the signs constituting the 'data' of communication-free observations can only be experienced as causal connections of natural processes, and can never be regarded as the communicative experiences of symbols expressing the intentions of Nature.

In contrast to this situation, the Human Sciences cannot restrict themselves to imposing hypothetical regularities (regarded as 'rules') onto their objects from outside and to testing these hypotheses by communication-free observation; for we cannot infer from an observed correspondence between a regularity imposed from outside and observed behaviour that the regularity is a 'rule' followed by the objects as subjects of actions. The 'rules' that operate in a society have to be understood (in the first instance) in terms of what it is like to be subject to such 'rules'. Also it is not possible to infer from observed deviations of human behaviour from supposed regularities, that these regularities express a valid rule (usually followed), or even a valid norm (although it is often followed). Clearly from this perspective the theses of the Human Sciences are not to be 'tested' in terms of the explanation of causal necessity in the manner of the Natural Sciences. Rather the focus of interest is on the reasons that can be postulated for understanding the meanings of intentions, motives, and actions of human subjects.

C) Time

If "... limited prognosis and causal explanation are different expressions for the same achievement ..." of the Natural Sciences ([HA4] p.137), then the reduction of the logic of

explanation to that of prognosis highlights the importance of the concept of 'time'. For the human subject, 'time' is a one-way arrow and is construed in terms of irreversible individual biography or social history. However as Apel puts it ([AP2] p.420), the Natural Sciences (i.e. physics)

"... need not deal with the world as history in a proper sense. By this I do not mean that there is no dimension of irreversibility, and hence of a history of nature, to be dealt with by physics. It is true, I think, that physics has to deal with irreversibility in the sense of the second principle of thermodynamics, i.e. in the sense of the increase of entropy. But, in this very sense of irreversibility, physics may suppose nature's being definitely determined concerning its future and thus having no history in a sense that would resist nomological objectification." (emphasis added).

In contrast, the Human Sciences cannot evade presupposing history

"... in a structural sense that resists nomological objectification" (p.20). For the Human Sciences "... must not only suppose irreversibility - in the sense of a statistically determined process - but irreversibility, in the sense of advance of human knowledge influencing the process of history in an irreversible manner" ([AP2] p.20 emphasis added).

Merton's famous discussion about self-fulfilling (and destroying) prophecies ([MT] pp.421-36) indicates that conditional predictions are not possible with respect to history. For the point of Merton's 'theorem' "... seems to lie in systematically precluding or preventing system isolation - in the very fundamental sense of a separation of the object-system and the subject-system of knowledge - by the understanding and reactive self-application of conditional predictions by the human subject-objects of those predictions" ([AP2] p.21). The destruction or fulfilment of conditional predictions (or prophecies) by human subjects might usefully be called 'Merton-effects'. Although communication-free

observation might influence natural phenomena when observation is also instrumental action, there are no Merton-effects in the study of natural phenomena. The complete elimination of the 'cognitive realisation and reactive self-application' of conditional predictions by human subjects (i.e. Merton-effects) in the Human Sciences would require the system isolation of experimentors and experimented (manipulators/manipulated) to ensure the reliability of conditional predictions. The difficulties involved in conceiving of such studies along the lines of communication-free Natural Science would seem to emphasise the differences in constitutive questions raised by the differing leading knowledge interests. On the one hand, the Human Sciences require a notion of historical time in which to understand intentions and motives of human subjects. On the other hand, the Natural Sciences require time to parameterise changes in natural phenomena.

D) Space

There could not be an experience that is not in time, and time is made real to us through the organisation of experience in terms of ordering of those changes called events. Similarly, nothing can appear to us as an independent object without being experienced as 'outside' and spatially related. Space, like time, forms part of the organisation of our sensibilities. Our sense-impressions bear the form of space, as is evidenced in the phenomena in the 'visual field'. Objects are referenced in everyday life by location in that field.

Our sense of that 'visual field' (or even geographical space), in which we move and have our being, can form a secondary

domain to ground an explanandum in terms of the metaphor, SPACE IS A CONTAINER. For example, in the context of a practical interest, a notion of 'social space' as a heirarchically-ordered social structure gains its meaning from a simple conception of society as a spatial entity containing class strata varying from regions of the 'upper' to regions of the 'lower'. On the other hand, the same metaphor is used in the context of a technical interest to ground causal explanations of the 'world of mechanical corpuscular motions'. The motion of a body takes place in space, for space lays down the foundations for the empirical description of motion. However, there is a crucial difference in the deployment of the metaphor with respect to the practical and technical interests. On the one hand, the technical interest of a Natural Science must give a certain priority to the schema 'space' if it is to use mathematics as a methodical device; for in order to describe the laws of Nature in terms of a causal connection, it is necessary to conceive 'substance' (moving body) and 'time' as spaces. On the other hand, the practical interest of the Human Sciences involves no such subsumption of its schema to that of 'space'.

The argument that the mathematical description of natural phenomena requires a notion of space as a container can be illustrated as below. In what follows the mathematical reader may assume the representation of space-time that suits him best. My illustration does not depend on any particular representation; the simplest Newtonian description is quite adequate. Let B denote a moving body, such a a system of 0-dimensional particles, a 1-dimensional cord, a 2-dimensional flexible shell, or a 3-dimensional solid or fluid body ([LWP] p.379). Let T denote time. Then T becomes a 1-dimensional space the moment it is conceived as

a continuum of points (such as R say) to reference instants of time. The body B also becomes a space, the moment we parameterise its particles by points of a space X (say the Euclidean Space R^3), and refer to its position in space by an embedding $B \rightarrow X$. A motion of the body in space is empirically described as a morphism

$$q: B \times T \rightarrow X,$$

which can be thought of as assigning to each particle in the body at each time instant its corresponding point in X. If the set of morphisms $T \rightarrow X$ is conceived as the set of paths in X, then space-time, denoted by X^T , can be regarded as the set of paths endowed with an appropriate topology identifying the cohesion of space-time. A morphism $\hat{q}: B \rightarrow X^T$ might be thought of as assigning a path in space-time to a body B. All of these objects, denoted by X, T, B, and X^T are spaces in so far as the mathematical physicist is concerned. The laws of Nature are described in terms of functional dependence of elements of these spaces, in which the space X is naturally thought of in terms of the explanans CONTAINER. Thus the gist of the Newtonian description, of a 'world of mechanical corpuscular motion', is that motion is change of location in space. Space is the container in which we can make sense of the motions of bodies in it. To determine the motions of bodies is to determine how they move in space.

Fine notes ([FI] p.449) that Newton's Scholium

"... ends with the words 'But how we are to obtain the true motions ... shall be explained more at large in the following treatise. For this end it was that I composed it'. And surely the implication is that space is just what physics says it is. The idea, I take it, is that a comprehensive physics is written in the language of space and time. There is a metric space-time geometry that underwrites physics in a way that inseparably entwines the two. Questions about space, then, become questions about this geometry and these

questions ultimately turn on what we take to be the true physical theory." (emphasis added).

Nevertheless, not everyone was to agree with Newton. To view space as a container is not necessarily mysterious, but the move is somewhat delicate. As Gadamer puts it "... space only becomes an object of thought by mentally removing the objects that are related to one another in it ..." ([GA1] p.392). Leibniz was not prepared to make this move. For him, space was not an unobservable container in which we make sense of our observations. Our experience of space is mediated as a concomitant of our experience of objects. He argued that space was a system (perhaps a lattice or network) of spatial relations. Fine notes that this relational view ([FI] p.448)

"... takes for granted the existence of bodies that bear to one another certain qualitative 'spatial' relations. Some bodies are shorter or longer than others, some are nearer or more distant than others, some are straighter or more bent than others. The difficulty for the relational view is to quantify this network of qualitative relations, for Newton (= Clarke) puts it to Leibniz, 'space and time are quantities; which situation and order are not'. Leibniz's response is to say that order also has its quantity and his effort to implement this remark constitutes a first rudimentary theory of measurement, of the sort associated with Helmholtz, Campbell, and their more recent descendents. The basic idea is to take some physical body as a standard for the relation in question and to specify a regimen for deploying this standard so as to construct a quantitative scale the ordering of which coincides with that of the given qualitative relation. In the case of spatial relations the various scales are forged together to form a metric geometry. Questions about space, then, become questions about this geometry and these questions ultimately turn on the deployment of physical standards." (emphasis added).

What do these differences between Newtonians and Leibnizians regarding the relation between geometry and physics mean for the Natural Sciences ? Does the Leibnizian rejection of the concept of space have a serious role to play in the Natural Sciences ?

Over the centuries, the controversy between Newtonians and Leibnizians has been somewhat ill-tempered and bitter. "Each side

has accused the other of being incoherent and, therefore, incomprehensible. And these charges have been looked upon by each side as a knock-out blow to the other" ([FI] p.478). Traditionally, the controversy has been conducted in terms of fundamental ontology. 'What is space?'. 'Is it something or nothing?'. 'If it is something, what sort of thing is it?'. Now Fine argues that this controversy lacks fundamental ontic bite, in that the ontological commitments of both viewpoints do not seem to cut deeply enough. Instead, he adopts a metascientific viewpoint and argues that the real differences between the camps revolve around contrasting research programs for approaching questions of space and time.

On the one hand ([FI] p.450),

"The Leibnizian program suggests that within the range of legitimate questions about the physical universe, there is an autonomous sphere of inquiry that has to do with questions about space and time. These questions assume a quantitative form and the research required to answer these questions concerns the possible deployment of physical bodies as standards of measurement. To carry on this research one may certainly employ auxiliary physical laws, but the import of the investigation is geometric (i.e. has to do with space and time). Ideally, the geometric results should be cleanly separable from the auxiliary research aids. (emphasis added).

On the other hand,

"... the Newtonian program is founded on a more holistic conception of the realm of legitimate inquiry ... that realm is relative to and completely circumscribed by physics itself. The research required to answer questions about space and time is research in physics. It concerns either the further development of physical theory or the investigation of the geometric commitments of current theory. The results of this research are part and parcel of physics. There is no clean separation between physics and geometry". (emphasis added).

Thus, Fine's attempt at conflict resolution reveals that, whatever merits both programs may have, the real differences revolve around the question of the separability of geometry and the laws of physics.

The historical merits of the two programs are brought out in the following ([FI] p.450) :

"Most notable among those who have carried on the program set by Newton are surely Poincaré and Einstein. Indeed the general theory of relativity certainly marks a high point in the historical development of the Newtonian scheme. Those who have followed the program set by Leibniz include important members of the school of Logical Positivism; notably, Schlick, Reichenbach and Carnap."

The Newtonian program seems to have been the one actually used by the Natural Sciences, whereas the Leibnizian program has mainly been the vehicle of philosophical research.

Whatever the merits of Leibniz's program for research into physical standards, the reasons why natural scientists have mainly worked with Newton's program are not difficult to find. Any serious attempt to use mathematics as a methodical device involves representing 'substance' and 'time' as spaces, and space-time as a container in which natural objects are embedded (or contained). It is unlikely that we could extricate our sense of space as geometry from the laws of physics as empirical descriptions of Nature. Fine argues that Einstein objected to the separability thesis ([FI] p.475), for

"... in order to carry out an empirical procedure for determining geometry it is necessary to correct our measuring instruments (rods etc.) for deforming influences. These corrections are made by means of physical laws, like the law of thermal expansion. These laws, however, are stated in quantitative geometric terms. They thus suppose the applicability of geometry. It follows that the most our empirical procedure could yield would be a conclusion of the form, 'If such-and-such geometry is assumed together with so-and-so physical laws, then such-and-so geometry is found to obtain'. It appears that the geometrical conclusion could never be detached."

Basically, this argument seems to say that the laws of nature (when formalised as natural invariances) contain statements of an intrinsic geometric character, which when suitably interpreted in the Conceptual yield the desired geometry. This follows from the

necessity of conceiving bodies (i.e. measuring rods) in geometric terms. The functional dependence of natural laws are only described in terms of geometric elements. From this perspective, the separability thesis looks untenable.

From the Newtonian viewpoint, space as a coordinatised framework parameterises space-time events. It is well-known that abstract sets can parameterise (or index) sets of 'things'. However, the subsets of a set are not quite like the parts of a space. True, the set of points of a space is a set. But there is more to a space than its set of points. Objects are referenced in space by spatial position, and in terms of distance and direction across paths and intervals. The points only serve their location-indexing function when they are regarded as the (end) points of a path across which the points are at a distance. Points and paths are particulars and their ability to spatially parameterise objects comes from the sort of cohesion (or topology) that a space has (NER] pp.16-28). Sets lack this cohesion. Thus it is the remarkable cohesive character of spaces (and not their set-like qualities) that is put to use in empirical descriptions of natural phenomena.

This brings me to my *donnée*. The constitutive questions of a technical cognitive interest stem from a schematization of the concepts of space, time, substance, and causality, in which natural phenomena are construed as non-intentional objects whose processes are to be explained in terms of a causal nexus. But the reduction of the logic of explanation to that of prediction requires the introduction of a geometric framework in which time and space coordinatise objectified processes. But this, in turn, requires

that natural invariances be described geometrically in terms of elements of spaces. From this, it follows that what is required from Mathematics, in terms of the efficiency of geometry as a methodical device, is a supply of 'spaces' appropriate to the research questions in hand. These 'spaces' might be thought of simply as 'objects' with 'geometric figures' in them. By the term 'geometric figure', I mean such geometric entities as 'points', 'paths', and 'infinitesimals', which cohere (by means of their topological organisation) into the object called a 'space'. What has yet to be established is that these objects live in topoi of spaces. Furthermore, the argument so far has been conducted entirely in terms of mathematics as a methodical device well-suited to the constitutive questions of the Natural Sciences. What has to be assessed now is the demands made on Mathematics by the Human Sciences. Do they need any mathematical conception of 'space' ?

CHAPTER FOUR

QUASI-CAUSAL AND STRUCTURAL REQUIREMENTS FOR SPACE

GENERATED BY THE HUMAN SCIENCES

I claim that the requirements for a supply of those objects called spaces from Mathematics arise directly out of the needs of Science. In the case of the Natural Sciences, the constitutive questions are primarily determined by a leading knowledge interest in instrumental action on Nature. The significance of mathematics as a methodical device is based on an approach to Nature, in which law-like hypotheses are framed in the spatio-temporally schematized imagination as a technical intervention which obliges Nature to answer our questions regarding a causal nexus ([AP3] p.47). Our inability to separate the laws of Nature from their geometrical expression serves to underline the significance of mathematics for the Natural Sciences. It is now hardly possible to conceive of a Natural Science in which natural events are not already imagined in terms of geometrical configurations. The requirements for a mathematical conception of space are obvious.

But what of the Human Sciences ? Do they need any conceptions of space from Mathematics ? There are certain obvious difficulties in answering this question. In the first place, if the leading knowledge interest of the Human Sciences is in the practical matter of understanding meaning and improving human communication, then the constitutive questions of the Human Sciences would preclude the involvement of Mathematics in anything like the same way as the Natural Sciences. From the perspective of

the theory of human cognitive interests, it would, in the language of Ryle ([RI] p.197), be tantamount to a category mistake to correlate the leading knowledge interest of these Sciences with a technical cognitive interest. This rules out an argument for a demand for spaces by the Human Sciences, as was constructed for the case of the Natural Sciences, by schematizing the concepts of space, time, substance, and causality in a manner conducive to the constitutive questions of a technical cognitive interest. In the second place, the wide scope of a practical cognitive interest means that the Human Sciences are both

- a) more ambitious and less efficient as explanatory knowledge than the Natural Sciences, in which the interest in knowledge is limited to that of instrumental action; and
- b) more amorphous and less homogeneous than the Natural Sciences, so that the intimate relationship existing between the Natural Sciences and Mathematics is unlikely to arise with the Human Sciences.

Attempts to portray as 'prescientific' those areas of the Human Sciences which resist colonization by Mathematics would be regarded by many of those working in the Human Sciences as a category mistake.

Nevertheless, I claim that the Human Sciences can generate a requirement for spaces from Mathematics. Furthermore, I claim that the conceptions of space required by the Human Sciences are essentially no different from those required by the Natural Sciences, in that spaces are conceived as objects in which geometric figures cohere. I base these claims on the possible second-order methodological objectifications which arise from the

complementarity of human knowledge interests. Although the practical interest is concerned with understanding meaning, that understanding is about some objective matter in the world. In so far as possible objectifications of the questions faced by the Human Sciences are amenable to metaphorical redescription in terms of a geometrical explanans, then these objectifications can ground the deployment of spatial concepts in the Human Sciences. However, one must proceed carefully here. If "... it is tempting to say that metaphor is a planned category mistake" ([RI] p.197), then some care is required in the controlled use of metaphor to plan that mistake, if we are to avoid an actual mistake. Accordingly, my objective in this Chapter is to show that, in the light of the complementarity thesis, a controlled use of metaphor (as a planned category mistake) can promote the demand for geometric objects in two possible ways :

- A) Just as the metaphor NATURAL PHENOMENON IS MOVING BODY permits the complete objectification of Nature for Sciences incorporating a technical cognitive interest, so the homology SOCIAL PHENOMENON IS TO NATURAL PHENOMENON AS SOCIAL PROCESS IS TO NATURAL PROCESS can metaphorically restructure some of the questions of the Human Sciences. I shall dub those Human Sciences, operating as if they incorporated a technical cognitive interest (as a planned category mistake), as quasi-causal (or quasi-nomological) Sciences. An example of a quasi-causal Human Science with a substantial mathematical content might be Econometrics.
- B) Even those Human Sciences which reject a quasi-causal approach to their subject matter may find it necessary to theorise about their questions in terms of structure. The homology STRUCTURE IS TO SPACE AS ORGANISATION IS TO COHESION can metaphorically

redescribe a sense of structure as an organised whole of connected parts in terms of an explanans drawn from the cohesive qualities of space. A concrete example might be theorising about social structure in terms of networks of social relationships conceptualised as a graph. Here the space is a graph in which the elements (nodes and edges) may cohere to form a whole of connected subgraphs. Now I have argued that space lays down the basis of the study of motion. But for those Human Sciences which do not proceed in a quasi-causal fashion, what is required is not so much a conception of space as a theatre of action, but as a cohesive object with differing regions in which to organise our experience of the objects under discussion. I shall call this a demand, for want of a better term, for a structural mathematics.

However, the metaphorical redescription of natural processes as objectified processes was disciplined within a framework which renounced any teleological understanding of Nature. If metaphor is planned category mistake, then the care involved in the planning rests on the renunciation of an intentional object. The demand for space from a Science incorporating a technical cognitive interest is made within the discipline of this framework. Now Apel argues that the complementarity of knowledge interests yield criteria of methodological differences between the various types of Human Science in spite of the apparent unity of the intentional subject studied by such Sciences ([AP2] pp.25-36). Broadly speaking, these methodological differences can be explored along two 'dimensions' :

1) QC/GRA dimension

The contrast between quasi-causal Sciences (QC) and the central

task of a practical cognitive interest, namely that of providing a good reason assay (GRA) for understanding the motives of the intentional subjects.

2) CN/NE dimension

The contrast between those Human Sciences which

- a) evaluate human actions and institutions in the light of those rules or norms that are constitutive for the actions or institutions as describable facts of the socio-historical world (CN-Sciences), and
- b) evaluate the very norms or rules by which the actions or institutions are constituted (NE-Sciences).

The distinctions made by Apel can be used to provide a simple framework of ideal types of Human Science. This framework can serve as a background against which to evaluate the demands for space generated by the Human Sciences. The sort of framework I have in mind is illustrated in Figure 3.

Apel's Ideal Types

	CN	NE
GRA	*	*
QC	*	*

Figure 3

The purpose of this framework is to clarify the elements giving rise to methodological differences in the Human Sciences and to set these out as ideal types so that when we approach a particular case of Human Science we shall know what features are especially worthy of consideration. In the remainder of this Chapter, I shall

- 1) outline some aspects of Apel's ideal types of Human Science, in so far as they form a useful backcloth to demands for space; and
- 2) show how Apel's framework may help us to assess the demands for spaces as objects in which geometric figures cohere.

1) Ideal Types of Human Science

The GRA/QC dimension marks a structural difference between those Human Sciences, such as the economic analysis, which proceed in a quasi-causal fashion, and those, such as History or Sociology, which try to understand the good reasons for socio-historical events. Now the interpretation of the presuppositions of people's mentality can never be taken for granted, as the history of the failure of economic projects in developing countries only too clearly demonstrates. Nevertheless, the results of the analysis of consumer behaviour may sometimes look like the results of Natural Science. Explanations and conditional predictions of consumer behaviour under certain situations are based on regularities 'observed' over a limited historical period. It is as if the analysis of consumer behaviour were based on a technical cognitive interest. Measured masses of goods are produced and consumed, and the movement of these measures through economic-space is the subject of limited predictions.

The famous ceteris paribus assumptions of economic analysis acknowledge the difficulties of making such predictions in the face of possible Merton effects, and point to the fact that such Sciences can never truly possess a technical cognitive interest. The regularities observed are not necessarily norms to be followed. However, these regularities are certainly the result of people

either conforming with or transgressing norms. These Sciences are dubbed quasi-causal with good reason; the as-if operator imports only a secondary methodological objectification. In the case of the Natural Sciences, the objectification of natural processes is disciplined within a schematization of substance as a non-intentional object. The causal nexus can be made the basis of if-then rules for successful instrumental action on Nature. However, metaphorical redescription of social processes (such as consumer behaviour) as objectified processes means that the application of derived if-then rules can only proceed quasi-causally by influencing a person's attitude or motives.

The concept of cause is clearly defined within a technical cognitive interest. However, the concept has its wider uses because we can continue to function with limited success when it is lifted from its natural situation in the technical cognitive interest into other areas through metaphorical redescription. For example, in 'Harry raised our morale by telling jokes', Harry moves our morale through morale-space by the input of telling jokes ([LJ] p.72). Certainly it seems sensible to use the term 'quasi-cause' in these instances when the application of a rule (whistle joyfully) 'causes' an effect (feeling cheerful) through the influencing of the intentional subject's attitude. The importance of quasi-causal Sciences lies in the application of the concept of quasi-cause to ever new domains of human activity. Universally quantified statements of if-then rules based on regularities described in quasi-causal terms can provide the basis for predictions under strong assumptions, such as ceteris paribus and the acknowledgement of Merton effects.

However, there are Human Sciences, such as History and Sociology, which cannot proceed in a quasi-causal fashion. Historians (or sociologists) attempt to offer explanations or good reason essays "... by which the actions of people can be understood and even justified as reasonable with respect to their aims and maxims, given the situational circumstances as they were understood by the actors themselves" ([AP2] p.26). These actions cannot be cast in in terms of quasi-causal if-then rules. Apel adduces the following reasons for this state of affairs ([AP2] pp.25-6) :

"... 'Because'-sentences of historical explanation, in contradistinction to 'because'-sentences of the quasi-nomological behavioural sciences cannot be reduced to nomological sentences of the form 'always if...then'. The reason for this impossibility is that historians, in contradistinction to behavioural scientists, are not allowed to refer their explanations by 'because'-sentences to quasi-laws and antecedent conditions of habitual social behaviour within the context of a social system that is itself still to be understood as relative to a certain period or region of history. Instead of presupposing such methodological abstractions, historians must explain historical events in the light of the whole of history to be considered as open to the future in principle."

If historians tried to proceed quasi-causally and offer nomological explanations of events, that is with universally quantified statements with respect to all possible historical events and all possible human reactions to events, then ([AP2] p.26) :

"... they would have to face the following dilemma : Either they would have to cautiously restrict their nomological premisses to propositions that would be so general and hence trivial that they would not be falsifiable and hence would not present relevant explanatory hypotheses; or they would have to provide nomological premisses that would contain definite descriptions and even proper names for all historical subjects (individual and collective) and for all their particular circumstances in life. In this case they would not have achieved a nomological explanation either, but at best they would have postulated the historical necessity, in principle, of the historical events, as in the proposition : 'In all cases where a ruler acts in the same way as Louis XIV did, and does so under the same conditions as the French king, he would necessarily lose his popularity at the end of his life, as Louis XIV in fact did'."

The difficulties faced by historians, in attempting quasi-causal explanations enable us to make a sharp distinction between quasi-

causal Sciences and those aiming at the provision of good reasons. To proceed quasi-causally (in relevant situations) is to demand more. It is to "... insist on ascertaining which intelligible (good or bad) reasons might be considered causally effective reasons ... as quasi-causes in the case of behaviour to be controlled by explanation and conditional prediction" ([AP2] p.27).

The CN/NE dimension is intended to structurally differentiate attitudes to value-laden descriptions in the Human Sciences. Now it is generally conceded that there can be no such thing as a value-free study of human institutions ([GO]). However, ([HA2] p.167) :

"... there can be different, yet coextensive, descriptions of the same event which are not synonomous, e.g. 'the death of Caesar' and 'the murder of Caesar'. However, the fact that Caesar was assassinated can only be rendered by one and the same statement. Coextensive but nonsynonomous propositions cannot express the same fact."

The meaning of murder is something that is defined by men. Whether or not Caesar was murdered is something to be intersubjectively agreed as a fact. Essentially, 'facts' are not happenings in the world ([HA2] p.169). For this reason the truth of propositions is not corroborated by objectified processes happening in the world, but through the intersubjective agreement achieved through argumentative reason. For as Habermas points out ([HA2] p.168) :

"When we say that facts are states of affairs which exist, we mean not the existence of objects, but the truth of propositional contents. In doing so we presuppose ... the existence of identifiable objects of which we state that they have such and such propositional content. Facts are derived from states of affairs, and states of affairs are the propositional content of statements, the truth claim of which is radically questioned. A state of affairs is the content of a proposition which is stated hypothetically rather than apodictically, i.e. the propositional content of a statement whose truth claim is held in abeyance. If a state of affairs is the content of a statement to the extent that it is problematic and gives rise to discourse, then what we call a fact is the content of a statement after it has been subjected to a

discourse that is now (for the time being) concluded. A fact is what we would want to assert as true after a discursive test. Facts are the contents of propositions where we 'maintain' what we had originally stated. In sum, the meaning of 'facts' and 'states of affairs' cannot be clarified without reference to 'discourses' examining the suspended truth claims of statements."

Now it seems to me that the basis of CN-Sciences lies in discourses which only evaluate what can be regarded as 'facts' in the light of norms or rules which are constitutive of those describable 'facts' of the social life-world. Although what it means for a proposition to be true is something established by human conventions, and even though (it is possible that) these conventions could have been otherwise, and again it may well indeed be true that these conventions are tainted with ideological components, it does not follow that a proposition cannot be objectively true. Thus we may decide what it means for a state of affairs to exist, but whether it is the case that a state of affairs is a 'fact' is something that can only be corroborated with the support of the objectivity of experience.

However, the suspension of validity-claims is double-edged, for it also seems that the basis of NE-Sciences lies in discourses which render problematic the norms or rules which constitute what a CN-Science would regard as a describable fact. For example, the hermeneutic clarification of Biblical texts within the context of Christian doctrine could be characterised as a GRA-CN Philological Science. On the other hand, as is well known, research into the historical Jesus by 19th Century Protestant theologians and pastors resulted in a Biblical Philology of a GRA-NE type in which almost all the elements of Christian doctrine were radically questioned. Now there is a case for the importance of both CN and NE type Sciences. In the case of a Philology of CN type, it may be

understood as a requirement for objectifying and presenting meanings that are to be grasped before judgements are made on the truth-claims of the text. But the significance of a Philology of NE type appears in situations in which progress in understanding meanings is not so much dependent on objective abstraction but on the competence of the investigators to arrive at a justification of the truth-claims of difficult texts. For example, Apel argues that the internal reconstruction of the History of Science as conceived by Lakatos would be impossible unless Science itself was placed on the basis of a continuing questioning of existing norms ([AP2] p.28). In a telling example ([AP2] p.32), Apel suggests that the Humanist philologists could not understand Archimedes' teachings on hydrostatics, until they critically reconstructed these teachings by separating out the justifiable modes of Archimedes' reasoning. To be sure, within the context of a GRA-Science, both CN and NE modes of procedure can coexist. Klein's reconstruction of the origins of Greek Mathematics involve both a CN approach, which improves our understanding of Mathematics in terms in which the Greeks understood it themselves, and an NE approach which reconstructs the history of Renaissance Mathematics in terms of a rupture with Greek Mathematics following a critical judgement regarding the wheat and chaff in Greek thinking ([KL]).

But what are the implications of the CN/NE dimension for QC-Sciences ? Since the QC-Sciences can never truly operate within the context of a technical cognitive interest, then this dimension can be used to characterise the relationship of QC-Sciences to possible Merton effects in the human domain of unrestricted communication.

On the one hand, there could be "... a steady fulfillment of

reciprocal expectations of behaviour by the process of human interaction" ([AP2] p.23). This case could be characterised as one of undisturbed cooperation by the human subjects, who are interested in the elimination of Merton effects by establishing the reciprocal fulfillment of behavioural expectations in accordance with predictions of the quasi-causal theory. The hope of undisturbed cooperation underlies the deployment of QC-Sciences as social technology, whether for market research or bureaucratic planning.

On the other hand, there could be "... a process of self-destroying prophecy, based on a constant thwarting and thereby frustrating of expectations" ([AP2] p.24). This case could be characterised as one of disturbed cooperation by the human subjects who capitalise on the unwanted predicted effects of quasi-causal theories as a situation to be avoided. The case for investment in QC-Sciences as the basis for thwarting self-fulfilling prophecies has been put by Apel ([AP1] p.57) :

"... these objectifications of certain aspects of human behaviour which cannot (yet) be articulated into the language of self-understanding nevertheless are serving to further this self-understanding. These attempts to let 'objectification' serve 'disobjectification', i.e. that condition in which man is freed by knowledge to act responsibly, have to be judged according to whether the 'objects' of the theory can become 'subjects' who can incorporate that theory into their own language and self-understanding".

This gives a first hint on how we may relate the CN/NE dimension to questions of cooperation with Merton effects, for the reactive self-application of the knowledge provided by QC-Sciences for disturbed cooperation by intentional subjects necessarily involves criticising the 'regularities' of quasi-causal laws as presupposing norms to be radically questioned.

I believe that we can understand this 'linkage' between regularities and norms, if we postulate the following equations :

- i) undisturbed cooperation = QC-CN type Science, and
- ii) disturbed cooperation = QC-NE type Science.

In the first case ([AP2] p.30),

"... it is fairly clear that abstention from evaluating rules as norms is required because the 'rules' here are interesting not as possible confirmations or alternatives to those rules we have to accept ourselves as obligatory norms for our life but only as supposed regularities or quasi-laws of a behaviour to be accounted for by quasi-nomological explanations. Hence abstention from value judgements is in this case a normative condition of the possibility of the object-constitution, precisely as in the case of the explanatory natural sciences."

Essentially the treatment of interesting regularities as quasi-laws involves suspending validity-claims as to the rightness of norms. Nevertheless these regularities are 'observed' in the context of norms which constitute the facts as facts. The undisturbed cooperation of human subjects with reciprocal expectations as to behaviour frozen in such quasi-laws must assume an expectation as to the continuation of norms and rules presupposed in such quasi-laws. But by the same token, a QC-NE type Science can treat the unwanted effects of a prediction based on these quasi-laws as components of an argument criticising the norms presupposed in such quasi-causal theories.

For those who reject Habermas' and Apel's propositions regarding an emancipatory cognitive interest (with a leading interest in knowledge which emancipates the human subject from coercion), an attractive possibility is to typify Sciences with such an emancipatory interest as QC-NE type Sciences. My own view is that the complementarity thesis rests on an idealisation which is necessary but not sufficient for an understanding of the full

scope of possible cognitive interests involved in studying the human subject. Actual social relations differ widely from the assumption of pure co-subjects debating academically the validity of rightness-claims in the esoteric realms of argumentative discourse. The compensation of the effects of manipulation by social technology requires a critical social science, which views itself as theory conceived as a practical-critical countervailing force. Nevertheless, the location of such critical social sciences in the QC-NE quadrant is sufficient for my purposes.

2) Space for the Human Sciences

The primary function of Apel's typification is to settle in our own minds where a particular Human Science should be located. But when we do this, it may be noticed that mathematics need not enter the Human Sciences. For example, in Table 3 I have indicated some examples of mathematical and non-mathematical types of Human Science.

Examples of Human Sciences

<u>Type</u>	<u>Mathematical</u>	<u>Non-Mathematical</u>
GRA-CN	Network Sociology	History/Sociology
GRA-NE	Network Sociology?	History/Sociology
QC-CN	Consumer Behaviour	Behavioural Psychology
QC-NE	Critique of Political Economy	Freudian Analysis

Table 3

It is unsurprising that mathematics need not enter the Human Sciences, for the theory of cognitive interests claims that the leading interest of such Sciences is in the practical questions of

improving human understanding; and whereas mathematics is a methodical device intimately bound up with the technical cognitive interest, mathematics only complements the practical interest as a secondary methodological abstraction when matters of objective description arise. I shall now restrict my discussion to those Human Sciences which import mathematics as secondary methodical device.

The 'dimensions' of Apel's typification can be used to clarify the demands made on Mathematics for conceptions of space. In the first place, the sharp distinction between QC- and GRA-Sciences can be used to show that the demand for space generated by the QC-Sciences is derived from that of the Natural Sciences, whereas the demand for a structural mathematics is only generated by a GRA-Science. In the second place, the CN/NE dimension can clarify the possible studies of processes parameterised by a spatial object. I shall deal with these in turn.

A) QC-Sciences

The point of Apel's typification of QC-Sciences is that in many respects they look like the Natural Sciences. The as-if operator enables us to import mathematics as a secondary methodical device in ways which are similar to the Natural Sciences.

Metaphorical redescription of social processes as objectified processes means that social phenomena are viewed as objects moving through a space. For example, an economic system may be viewed as a vehicle with a trajectory through an economic space, in which the coordinates (or points) are parameterised by quantities such as prices, wages, profit rates, time discount rates, and rates of

interest. But already in this example, we see that the demand for space is really no different from that of the Natural Sciences. Space is required to support the interest in 'motion', and this lays down a demand for objects with such geometric figures as 'points' to mark the position of the social system, and 'paths' or 'trajectories' along which a system may 'move'. The demand for space is essentially derived from that of the Natural Sciences, where the derivation is justified by grounding the description of social processes in the concept of (quasi-)cause.

B) GRA-Sciences

Although the GRA-Sciences deliberately eschew quasi-causal methods, they cannot avoid reference to objective matters such as social structure. For example in Chapter 2, I referred to Habermas' discussion of the practical cognitive interest. There, Habermas found it necessary to characterise the transformation of the neonate into a social individual as one of "... entrance into a network of communicative relations ..." ([MC] p.56, emphasis added). But what is social structure that it can be referenced by 'networks' ? A useful approach to answering this question is provided by Gould et al. ([GJC] pp.S5-S6) :

"It would be our contention that ... such knowledge [of structure] consists of how different things are connected together, and what these collections of things mean to us. In essence, then, much of our knowledge must be structural, precisely because structures consist of things connected together in certain ways ... it is hardly inappropriate to search for an effective way of describing the structures of the human world ... Now if we are going to talk about structure in well-defined and operational ways, and so to make it a concept that can be genuinely useful in creating agreed-upon and verifiable knowledge ... then we must try to think through what sort of 'language' is actually appropriate for theoretical expression and empirical description. And during the course of our thinking, one of the things we must try to avoid is the uncritical borrowing of mathematical languages that have been developed ... [for] ... the physical world ... So we are going to focus upon the question of structure ... the way things are connected and hang together ... this means we have to talk about sets and relations."

Now it seems to me that there are three important cross-currents in the above discussion

- i) the idea that structure has meaning and can be discussed in a precise and lucid fashion;
- ii) the idea that structure is a connected something-or-other; and
- iii) the idea that structure can be referred to in a precise way through a mathematical language of sets and relations.

If we take up these hints, I suggest that we think of structure in terms of the 'sum',

$$\text{STRUCTURE} = \text{LOGIC} + \text{COHESION}.$$

Firstly, the idea of connectivity is essentially a 'spatial' idea. Connectivity is but an element in our sense of cohesion. The most primitive idea of 'space' is a 'network'. Here nodes may act as coordinates to model social location or position in 'social space', and arcs might describe social relations which connect the nodes to provide a picture of social structure. In fact "... the language of graphs can be used to model many notions that were current long before graph theory was born" ([BH] p.236). For the mathematician, a graph is the simplest form of space, in which geometric figures of vertices (nodes) and edges (arcs) cohere (are connected) to form a whole or collection of wholes. Secondly, it is natural for a mathematician to generalise our notions of sets and relations to objects and inclusions. Just as a binary relation on a set A can be described as a subset R contained in the set $A \times A$, so any n -ary predicate can be described as an 'inclusion' of an object R in an object A^n in a set-like category ([GB] pp.239-48). For a precise description of structure, a mathematician will require a formal quantified statement expressed as an n -ary predicate in some formal

mathematical language, for which the cohesive object (i.e space / graph) is a model. For example, Seidman and Foster discuss how Menger's Theorem might facilitate an analysis of generalised cliques in the study of closely-knit human groups ([SF1]). A generalised clique is a model of some theory stated in precise axiomatic form capturing precise relationships between the elements of a graph. The quest is for precise descriptions of social structure. Thus the discussion of structure revolves around the two components of the 'sum'. On the one hand, the structure must demand an object (such as a graph) in which geometric figures (such as vertices and edges) cohere. On the other hand, the precise clarification of structure, as an expression of relations existing between geometric figures, requires a logical mathematical language.

Within the context of GRA type Sciences, there is a demand for a structural mathematics to describe in precise terms just exactly what structures are being referred to. Mathematics is but an ingredient in the good reason assay provided for understanding a given social situation. Mathematics does not provide the meaning of the situation, but it helps, in terms of a precise definition of a referenced structure, to support any meanings we may arrive at. Now I have argued that space supports the idea of 'motion'. What is required in a description of structure is not so much a support for 'motion', but rather support for a cohesive structure to 'coordinatise' the meaningful reference to human subjects. This is what is required for an idea of 'social space' - the identification of subregions to be roughly correlated with identifiable human behaviour or socio-historical events. To be sure, even in a simple space such as a graph, the possibilities of 'motion' are seen in

the support a graph can provide for 'paths' through it. And surely these geometric figures known as 'paths' in a space reference possibilities of communication between subregions. But the moment we reference communication in objectified terms as transmission of an 'object' along a path, then we are moving in the direction of a QC-Science. For this reason, I emphasise that it is the cohesive qualities of space that is required by a GRA-Science in its search for appropriate ways of talking about structure.

C) CN/NE Dimension

Whereas the QC/GRA dimension leads to sharp distinction enabling us to grasp that a GRA type Science requires only a structural mathematics, the CN/NE dimension only introduces the possibility of differing approaches to the study of processes. In the case of studies of morphogenesis, a Natural Science may need a space X in which to parameterise all possible configurations to express the state of a body B. What may be needed is a study of all the processes by which a space X may continuously index other spaces A, B, C etc. In my view because CN-Sciences only envisage undisturbed cooperation, there is an analogy with the Natural Sciences in that they may also require a study of X-parameterised processes, containing all the geometry compatible with the continuation of presupposed constitutive norms. However in contrast, the NE-Sciences, precisely because they envisage disturbed cooperation, may need to study all the ways in which a structure may be continuously transformed from X-indexed objects to Y-indexed objects. To be sure, from a mathematical perspective the study of X to Y parameterised processes may not be intrinsically different from the study of X- or Y-parameterised processes, if the

processes from X to Y can be referenced by a space Y^X . Even so, what may be required for Science is a mathematical conceptualisation of spatial processes which permits this leap to the study of higher-order processes.

Thus I claim that the variation in demand for space by the Human Sciences can be adequately accounted for by the theory of cognitive interests. The complementarity of knowledge interests permits a simple idealised typology, in which the demand for space can be viewed in terms of either the quasi-causal support for objectified processes, or the expression of the reference to social structure or social space. Similarly the possibility of typing Human Sciences in terms of differences in questions of value-laden description may lead to a demand for the study of parameterised processes whether simple or higher-order. Whatever else this pattern of demand for space is, it is always a demand for objects in which geometric figures cohere. I have used the term 'space' as a shorthand for this, and really this needs some justification. My argument is that conceptions of space required by the Human Sciences is essentially no different from that required by the Natural Sciences. Now it may be readily appreciated that the demand for space by the QC-Sciences is on the same footing as that of the Natural Sciences; basically quasi-causal laws require a conception of space to support an objectified description of 'motion' of social processes. But in what sense is a demand for a structural mathematics a demand for space which is essentially the same? What is the essence here? Now there is a well-known technique in Mathematics by which one can view apparently different things as abstractly the same - that is the use of category theory, in which the operational complexities of widely differing branches of

mathematics are put on the same footing, in terms of objects and arrows, to exhibit their conceptual similarity. To see that these conceptions of space are essentially the same, we will have to take a mathematical tack through the categorical treatment of set-like categories.

CHAPTER FIVE

SOME CATEGORICAL CONCEPTS

It is a commonplace that categories are there to make the different topics of mathematics more transparent by revealing common underlying patterns. For a category theorist, it is "... not Substance but invariant Form that is the carrier of relevant mathematical information" ([LWS] p.1506). For example, Goldblatt argues that ([GB] pp.1-2) :

"A category may be thought of in the first instance as a universe for a particular kind of mathematical discourse. Such a universe is determined by specifying a certain kind of 'object', and a certain kind of 'arrow' that links different objects. Thus the study of topology takes place in a universe of discourse (category) with topological spaces as objects and continuous functions as arrows ... and so on. We may thus regard the broad mathematical spectrum as being blocked out into a number of 'subject matters' or categories (a useful way of lending coherence and unity to an ever proliferating and diversifying discipline). Category theory provides the language for dealing with these domains ..."

Thus one way of looking at category theory is to see it as classifying various branches of mathematics with respect to underlying common patterns. However, "... category theory does not rest content with mere classification ... (although a few of its practitioners may do so); rather it is the mutability of mathematically precise structures (by morphisms) which is the content of category theory ..." ([LWY] p.148). Thus Lawvere notes that the development by category theory of methods of passing from one domain to another has had the effect of altering (advanced) mathematical practice itself ([LWY]). 'Such a universal instrument for guiding the learning, development, and use of advanced mathematics cannot fail to have its implications' for mathematical practice. If Science is the rational reorganisation of knowledge

achieved through a practice that alters its present form, then the altered practice of mathematics has implications for Science (insofar as Mathematics is a component of the Sciences). Thus category theory acquires a metascientific flavour when it alters mathematical/scientific practice. Category theory has come to be seen as the reflection of Mathematics on itself. It is with good reason that Lawvere describes category theory as a form of objective dialectics ([LWI]).

This Chapter is not intended to provide a comprehensive introduction to category theory for the non-mathematician; such an introduction would be a formidable undertaking and beyond the scope of a thesis of this character. My aim is to outline the (naive) category theory ([BN]) I shall regard as familiar, by giving the definitions, results, and notation I shall be assuming in later chapters.

The three fundamental concepts of category theory are categories, functors, and natural transformations.

Definition 1 A category C consists of three things :

- a) A class of objects, usually denoted by capital letters A, B, C, \dots
- b) A class of arrows or morphisms, usually denoted by lower-case letters f, g, h, \dots . Each morphism has a domain and codomain which are objects of C; we write ' $f: A \longrightarrow B$ ' for ' f is a morphism and $\text{dom}(f) = A$ and $\text{cod}(f) = B$ '.
- c) A composition law which assigns to each pair of morphisms (f, g) with $\text{dom}(f) = \text{cod}(g)$ a composite morphism $f \cdot g: \text{dom}(g) \longrightarrow \text{cod}(f)$. (Sometimes we shall just write fg or $f \circ g$ for the composite arrow). Composition is required to satisfy two axioms :
 - (i) Composition is associative, i.e. $(fg)h = f(gh)$ whenever

composites are defined.

(ii) For each object A , there exists an identity morphism

$\text{id}_A : A \rightarrow A$ satisfying $f \cdot \text{id}_A = f$ and $\text{id}_A \cdot g = g$, whenever we have arrows $f: A \rightarrow B$ and $g: B \rightarrow A$.

Definition 2 A functor $T: \underline{C} \rightarrow \underline{D}$ is a morphism of categories.

Specifically it consists of functions

(objects of \underline{C}) \rightarrow (objects of \underline{D}) and

(morphisms of \underline{C}) \rightarrow (morphisms of \underline{D}),

both denoted by T , such that

(i) If $f: A \rightarrow B$, then $T(f): T(A) \rightarrow T(B)$.

(ii) $T(fg) = T(f) \cdot T(g)$ whenever fg is defined.

(iii) $T(\text{id}_A) = \text{id}_{T(A)}$ for all A .

Definition 3 A natural transformation $t: S \rightarrow T$ between two

functors $S, T: \underline{C} \rightarrow \underline{D}$ consists of a function

(objects of \underline{C}) \rightarrow (morphisms of \underline{D}), denoted by $A \mapsto t(A)$,

such that

(i) $t(A): S(A) \rightarrow T(A)$ for all A .

(ii) For all $f: A \rightarrow B$ in \underline{C} , we have $T(f) \cdot t(A) =$

$t(B) \cdot S(f): S(A) \rightarrow T(B)$.

Examples of categories abound. Perhaps the paradigm category is the familiar category Set of sets and functions. Mathematicians frequently deal with concrete categories, whose objects are sets with some kind of structure and morphisms are the structure-preserving functions; the composition law being the usual composition of functions. Familiar categories are :

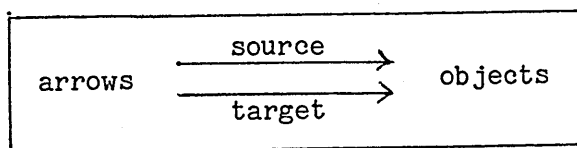
<u>Category</u>	<u>Objects</u>	<u>Arrows</u>
<u>Lat</u>	lattices	homomorphisms
<u>Grp</u>	groups	homomorphisms

<u>AbG</u>	abelian groups	homomorphisms
<u>Mon</u>	monoids	homomorphisms
<u>Ring</u>	rings	homomorphisms
<u>Bool</u>	Boolean algebras	homomorphisms
<u>Poset</u>	posets	monotone maps
<u>Met</u>	metric spaces	Lipschitz maps
<u>Sp</u>	topological spaces	continuous maps
<u>Man</u>	compact manifolds	smooth maps

Note that a concrete category C is always locally small. That is, for each pair of objects (A,B), the class of arrows $A \longrightarrow B$ in C (denoted by $\text{hom}_{\underline{C}}(A,B)$ or $\underline{C}(A,B)$) form a set rather than a proper class.

Useful categories for highlighting properties of topoi (of spaces) are certain categories of graphs.

Definition 4 A directed graph consists of two classes, the class of arrows and the class of objects, together with two mappings between them called source and target, as in



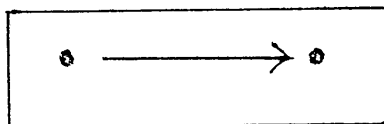
Instead of saying that $\text{source}(f) = A$ or $\text{target}(f) = B$, we write more briefly $f: A \longrightarrow B$ or $A \xrightarrow{f} B$, where f is an arrow, and A and B are objects. Sometimes the mappings source and target are called domain and codomain respectively. Borrowing from sheaf theory, we will call these source and target maps restriction maps. Sometimes arrows will be called edges, and objects called nodes or vertices. Edges whose source and target are the same node will be called loops.

A graph homomorphism from a graph G to a graph H is a function $V:G \rightarrow H$ sending nodes to nodes, and a function $E:G \rightarrow H$ sending edges to edges so that, whenever $f:A \rightarrow B$ is an edge in G , then $E(f):V(A) \rightarrow V(B)$ is an edge in H . A graph homomorphism preserves the source and target of an edge. That is

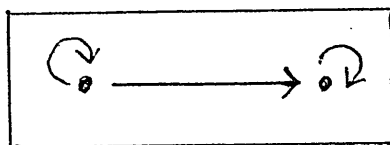
$$\text{source}(E(f)) = V(A) \text{ and } \text{target}(E(f)) = V(B).$$

Directed graphs and homomorphisms form the category Graph.

Definition 5 A reflexive directed graph is a directed graph in which every node has exactly one special loop known as the identity loop or degenerate loop. Of course, it may have other loops as well, but the degenerate loops are always there. A homomorphism of reflexive directed graphs is just a homomorphism of (irreflexive) directed graphs in which degenerate loops are always sent to degenerate loops. In a reflexive graph, we may picture the graph with one edge and two degenerate loops as :



with the degenerate loops suppressed. However, the same graph, when considered as an irreflexive graph must be pictured as :



in which all loops are shown. A homomorphism of irreflexive graphs is just a graph homomorphism. No special consideration is given to the question of the degeneracy of a loop. It merely asks that loops be sent to loops. Reflexive directed graphs and homomorphisms form the category RGraph. It can readily be seen that a category is a reflexive directed graph with an additional structure of

composition on the edges (Definition 1).

New categories can always be obtained from old.

Definition 6 Given categories \underline{C} and \underline{D} , we can construct the product category $\underline{C} \times \underline{D}$ in the obvious way : objects are pairs (C,D) with C an object of \underline{C} and D an object of \underline{D} . An arrow $f \times g: (C,D) \longrightarrow (C',D')$ has $f: C \longrightarrow C'$ in \underline{C} and $g: D \longrightarrow D'$ in \underline{D} .

Definition 7 If A is an object of \underline{C} , the slice category \underline{C}/A of objects of \underline{C} over A has as objects (B,f) all arrows $f: B \longrightarrow A$ of \underline{C} with target A . An arrow of \underline{C}/A is an arrow $h: B \longrightarrow C$ of \underline{C} making the following diagram commute :

$$\begin{array}{ccc} B & \xrightarrow{h} & C \\ & \searrow f & \swarrow g \\ & A & \end{array}$$

A typical slice category is $\underline{\text{Set}}/X$ in which the set X indexes other sets. Labelled graphs are objects in the category $\underline{\text{Graph}}/X$, in which (in the simplest case), the graph X has one node and a loop for each label. A labelled graph (A,f) is then just a graph A equipped with a graph morphism $f: A \longrightarrow X$ assigning a label to each edge of the graph A .

Categories may include subcategories.

Definition 8 A subcategory \underline{C} of a category \underline{B} is any category whose class of objects and arrows is contained in the class of objects and arrows in \underline{B} , and which is closed under the operations of source, target, identity, and composition. Subcategory \underline{C} is full when for any objects C, C' of \underline{C} , $\underline{C}(C,C') = \underline{B}(C,C')$.

Definition 9 A functor $F: \underline{A} \longrightarrow \underline{B}$ is faithful, if the induced mappings $\underline{A}(A,A') \longrightarrow \underline{B}(F(A),F(A'))$ sending $f: A \longrightarrow A'$ onto $F(f): F(A) \longrightarrow F(A')$ for all A, A' in \underline{A} are injective, and full if

they are surjective. A full embedding is a full and faithful functor which is also injective on objects, that is $F(A) = F(A')$ implies $A = A'$.

Arrows can be classified by their qualities.

Definition 10 A morphism $f:A \rightarrow B$ in \mathcal{C} is called

- a) invertible, if there exists an arrow $g:B \rightarrow A$ in \mathcal{C} with $gf = \text{id}_A$ and $fg = \text{id}_B$. Then we write $A \cong B$ and say that A and B are isomorphic and f is an isomorphism;
- b) monic (or mono, or monomorphism) if whenever $p:C \rightarrow A$ and $q:C \rightarrow A$ with $fp = fq$ then $p = q$;
- c) epic (or epi, or epimorphism) if whenever $p:B \rightarrow D$ and $q:B \rightarrow D$ with $pf = qf$ then $p = q$.

Definition 11 Given a morphism $f:A \rightarrow B$, a section is an arrow $g:B \rightarrow A$ with $fg = \text{id}_B$, and a retraction is an arrow $h:B \rightarrow A$ with $hf = \text{id}_A$. If f has a section g then f is epic. If f has a retraction h then f is monic. If g is a section then g is monic. A category in which every arrow is invertible is called a groupoid.

A close inspection of some categories reveals that their objects may themselves be categories.

Definition 12 A category \mathcal{C} is said to be small if the classes of objects and arrows are sets.

For example, a topological space X (in Sp) may be viewed as a category whose objects are the open sets of X , and whose arrows are the inclusion mappings between them. More generally, any poset may be regarded as a category : elements are objects, and there is at most one arrow $A \rightarrow B$ for any pair of elements whenever $A \leq B$. A monoid is a category with one object (the unit), and with elements as arrows. Even a set may be looked upon as a discrete category in

which there are no arrows other than identities.

From any category, we can form the dual category.

Definition 13 Given a category \underline{A} , the dual or opposite category, denoted by \underline{A}^{op} , has the same objects as \underline{A} , but for each morphism $f:A \rightarrow B$ in \underline{A} we get an opposite morphism $f^{op}:B \rightarrow A$ in \underline{A}^{op} , by interchanging the domain and codomain of f . Composition in \underline{A} is defined by $f^{op} \cdot g^{op} = (g \cdot f)^{op}$ whenever $g \cdot f$ is a composable pair of arrows in \underline{A} ([D01] p.8).

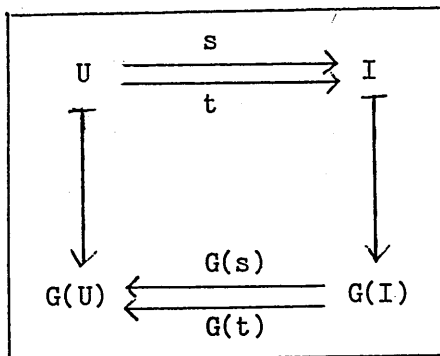
A functor from \underline{A}^{op} to \underline{B} is often called a contravariant functor from \underline{A} to \underline{B} . Functors which are not contravariant are sometimes said to be covariant.

The category of small categories and functors between them is called Cat. Here functors are treated as morphisms of Cat. A more startling idea is to treat functors from \underline{A} to \underline{B} as objects and natural transformations between them as morphisms in a functor category $\underline{B}^{\underline{A}}$. Functors from small categories into Set form the basis of our understanding of variable sets.

Definition 14 A variable set is a (contravariant) functor from a small category \underline{C} to Set. For each set-valued functor F , each object U in \underline{C} is called a stage of definition of F , and $F(U)$ is called the set of elements of F defined at stage U . For each $f:U \rightarrow V$ in \underline{C} , the elements of $F(U)$ and $F(V)$ are said to (contra-)vary along the restriction map f . In particular, a variable set of the form $\underline{C}^{op} \rightarrow \underline{Set}$ is called a presheaf.

For example, consider the category $\underline{P}(2)$ represented by $U \xrightleftharpoons[t]{s} I$. It has two objects and four arrows. A contravariant set-valued functor, $G:\underline{P}(2) \rightarrow \underline{Set}$ is a pair of sets $G(U)$ and $G(I)$, and a pair of arrows $G(s):G(I) \rightarrow G(U)$ and $G(t):G(I) \rightarrow G(U)$. The two identities are sent to identities. Such a functor G can be regarded as an irreflexive graph G . The elements of $G(U)$ are the nodes of G ,

and the elements of $G(I)$ are the edges of G . The edges, $G(I)$, contravary along the restriction maps $G(s)$ and $G(t)$ to the nodes, $G(U)$, and are said to restrict the edges to their source and target nodes respectively. We can represent this iconically as



Thus irreflexive graphs can be viewed as presheaves in $\text{Set}^{\mathcal{P}(2)^{op}}$.

Also a graph homomorphism $f:G \rightarrow G'$ is a natural transformation of functors (= graphs) in $\text{Set}^{\mathcal{P}(2)^{op}}$. In a similar vein, a set itself may be regarded as a functor from the discrete one-object category $\mathbf{1}$ to Set , where it can readily be seen that all variation is frozen into a constant object.

Other important classes of variable sets are the categories of \underline{M} -sets. If $(M, \cdot, 1)$ is a monoid seen as a one-object category with elements as arrows (Definition 11), then an \underline{M} -set may be regarded as a functor from \underline{M}^{op} to Set . Thus an \underline{M} -set is a set A with a map $A \times M \rightarrow A$ describing a right action by the monoid M on A . This is usually given in terms of $(a, m) \mapsto (a \cdot m)$, such that $a \cdot 1 = a$ and $(a \cdot (m \cdot m')) = ((a \cdot m) \cdot m')$ for all $a \in A$ and $m, m' \in M$. For example, consider the monoid $(M, \cdot, 1)$ with $M = \{1, s, t\}$ and the following binary operation :

\circ	1	s	t
1	1	s	t
s	s	s	s
t	t	t	t

The right action of M on a set A is equivalent to a reflexive graph. If M acts on A , then A is a set of edges of a reflexive

graph and a.s and a.t are degenerate loops associated with the source and target nodes of edge $a \in A$. Thus an M-set is a variable set with one stage of definition but in which variation may occur along differing arrows.

The central concept of category theory is the notion of adjunction or adjointness.

Definition 15 Given functors $F:\underline{A} \longrightarrow \underline{B}$ and $G:\underline{B} \longrightarrow \underline{A}$, we say F is left adjoint to G (denoted by $F \dashv G$) and G is right adjoint to F , if there is a bijection natural in the variables A and B between morphisms $f:A \longrightarrow G(B)$ in \underline{A} and morphisms $\bar{f}:F(A) \longrightarrow B$ in \underline{B} (See Fig.4), such that $\underline{B}(F(A),B) \cong \underline{A}(A,G(B))$ for all objects A in \underline{A} and B in \underline{B} .

Picture of Adjunction

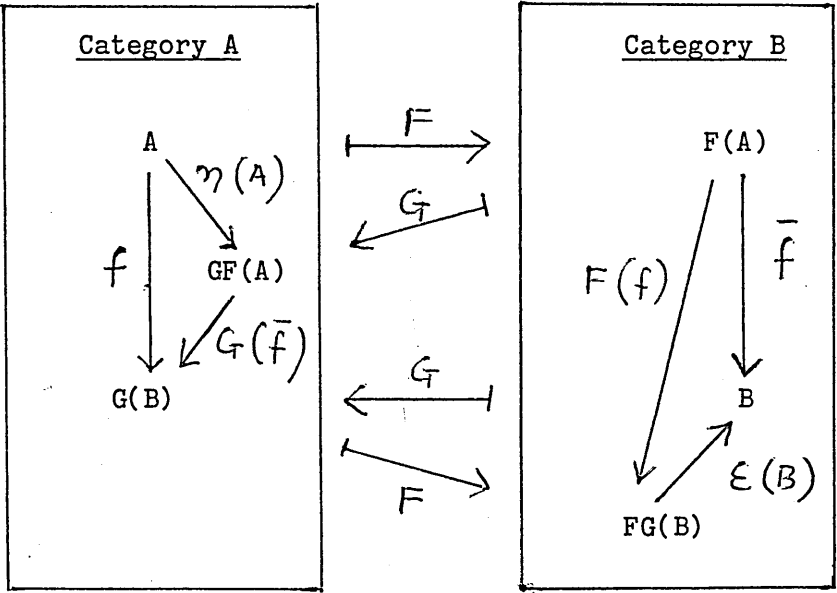


Figure 4

The adjoint relationship is often presented schematically by

$$\frac{A \longrightarrow G(B)}{F(A) \longrightarrow B}$$

which displays the left-right distinction. Given an adjunction $(F \dashv G)$, we may consider for each A in \underline{A} the arrow $\eta(A): A \dashrightarrow GF(A)$ which corresponds in the bijection to $\text{id}: F(A) \dashrightarrow F(A)$ in \underline{B} . Naturality of $f \dashrightarrow \bar{f}$ means that η is a natural transformation $\text{Id}_{\underline{A}} \dashrightarrow GF$ called the unit of the adjunction. Dually we can define a natural transformation $\epsilon: FG \dashrightarrow \text{Id}_{\underline{B}}$ called the counit of the adjunction.

Proposition 1 A functor $G: \underline{B} \dashrightarrow \underline{A}$ has a left adjoint provided we can find for each object A in \underline{A} an object $F(A)$ in \underline{B} and a morphism $\eta(A): A \dashrightarrow GF(A)$ which is universal among the morphisms from A to the image of G , in the sense that for any $f: A \dashrightarrow G(B)$ there is a unique $\bar{f}: F(A) \dashrightarrow B$ satisfying $\bar{f} = \epsilon(B) \circ F(f)$ with $f = G(\bar{f}) \circ \eta(A)$. (See Figure 4).

Proof : ([ML]) p.81).

The most familiar example of an adjunction is as follows. Let $U: \underline{\text{Grp}} \dashrightarrow \underline{\text{Set}}$ be the forgetful functor which sends a group to its carrier set and a homomorphism to its underlying function. If X is a set, the defining property of the free group $F(X)$, generated by X , says precisely that the inclusion $X \hookrightarrow F(X)$ is universal among functions from X to the image of U . So U has a left adjoint, the free functor $F: \underline{\text{Set}} \dashrightarrow \underline{\text{Grp}}$. For further examples of adjoints, see MacLane ([ML] p.85).

Adjunctions are the essential tool to explore similarities between categories.

Definition 16 Let $F: \underline{A} \dashrightarrow \underline{B}$ be a left adjoint for $G: \underline{B} \dashrightarrow \underline{A}$. A reflection is an adjunction for which the counit map $\epsilon(B)$ is an isomorphism for all B in \underline{B} . This is equivalent to saying that G is full and faithful and \underline{B} is a reflective subcategory of \underline{A} . That is, the adjunction induces a bijection between $\underline{B}(B, B')$ and

$\underline{A}(G(B), G(B'))$. If both the unit and counit are iso then \underline{A} and \underline{B} are said to be equivalent categories, denoted by $\underline{A} \simeq \underline{B}$.

This notion of equivalence is weaker than that of isomorphism of categories (denoted $\underline{A} = \underline{B}$), which may be regarded as an adjunction for which the unit and counit maps are identities. It can readily be seen that Graph and $\text{Set}^{P(\mathcal{A})^{op}}$ are equivalent categories. Also the right action of the monoid on a set of arrows (mentioned after Definition 14) is equivalent to RGraph. Nevertheless equivalence is sufficient to ensure that two categories share the same categorical properties. Indeed sharing categorical properties through categorical equivalence more or less defines what a categorical property is.

Adjoint functors can be used to define the important categorical concepts of limit and colimit.

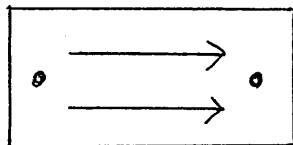
Definition 17 Let \underline{A} and \underline{J} be categories, with \underline{J} invariably small. Let $\underline{A}^{\underline{J}}$ be the functor category, whose objects are functors $D: \underline{J} \rightarrow \underline{A}$ (called in this context diagrams of type \underline{J} in \underline{A}) and whose morphisms are natural transformations. There is an obvious constancy functor $\Delta: \underline{A} \rightarrow \underline{A}^{\underline{J}}$, which sends an object A in \underline{A} to the constant functor in $\underline{A}^{\underline{J}}$ with value A . We say \underline{A} has limits of type \underline{J} if Δ has a right adjoint $\varprojlim_{\underline{J}}$. We refer to $\varprojlim_{\underline{J}}(D)$ as the limit of diagram D . Dually if Δ has a left adjoint (denoted by $\varinjlim_{\underline{J}}$), then \underline{A} has colimits of type \underline{J} . Limits and colimits are dual concepts. A colimit of $D: \underline{J} \rightarrow \underline{A}$ is just the limit of $D^{op}: \underline{J}^{op} \rightarrow \underline{A}^{op}$. ([ML] pp.62-71) (See Note 2, p.224).

I shall not give a detailed account of the various limits and colimits. They can be found in any standard text on category

theory ([ML]). The following examples outline important (co)limits in Set. For variable sets (co)limits are often constructed pointwise, so the reader should have little trouble in transferring his knowledge of (co)limits in Set to variable sets. Consider the following diagrams of type J.

- (a) Let J be the empty category 0, then a limit is a terminal object (denoted by 1), and a colimit is an initial object (denoted by 0). These objects are characterised by there being only one function $!_A : A \rightarrow 1$ and $0_A : 0 \rightarrow A$ for any object A. In Set, a terminal object is any singleton set $\{x\}$, whereas the initial object is the empty set \emptyset .
- (b) Let J be the category consisting of two discrete objects, then a limit is a categorical product of two objects (denoted by $A \times B$) and a colimit is the coproduct (denoted by $A + B$). In Set, a product of a pair of sets (A,B) is the familiar cartesian product $A \times B$, whereas the coproduct is the disjoint union of sets $A + B$. To be sure, products and coproducts can be generalised into n-ary products $(\prod_i A_i)$ and coproducts $(\sum_i A_i)$ by making J the category of n discrete objects. Products and coproducts have morphisms $p_i : A^n \rightarrow A$ (projection onto the ith factor) and $i_j : A_j \rightarrow \sum A$ (injection of the jth component).

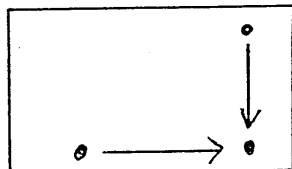
- (c) Let J be the category represented diagrammatically by



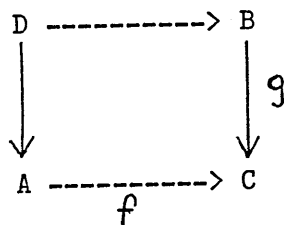
A limit of diagrams of type J is an equaliser. Dually a colimit is a coequaliser. In Set, an equaliser of a pair of parallel functions $f:A \rightarrow B$ and $g:A \rightarrow B$ is the inclusion of $E = \{a \mid f(a)=g(a)\}$ in A. Coequalisers are somewhat more complicated.

Let S be the binary relation $\{(f(a), g(a)) \mid a \in A\} \subseteq B \times B$. Then define R to be the finest equivalence relation on B containing S . The coequaliser is then given by the rule $b \in B \mapsto [b] \in B/R$, where the quotient set B/R is the codomain of the coequaliser ([GB] p.63).

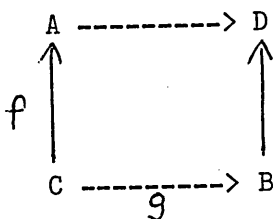
(d) Let \underline{J} be the category represented diagrammatically by



A limit of diagrams of type \underline{J} is called a pullback. Dually a colimit is a pushout. In Set, we consider diagrams like



in which $D = \{(a, b) \mid (a, b) \in A \times B \text{ and } f(a) = g(b)\}$. The set D is called a pullback (or fibred product), and is sometimes denoted by $A \times_C B$. Again pushouts are more complicated. Consider the diagram



with given functions f and g . We form the pushout D by defining the sum $A + B$ and then coequalising $i_A \cdot f: C \rightarrow A + B$ and $i_B \cdot g: C \rightarrow A + B$, and then defining D as the coequaliser.

I will not consider more complicated types of (co)limits, thanks to

Proposition 2 If a category has equalisers and all finite (resp. all small) products then it has all finite (resp. all small)

limits. Dually for colimits.

Proof : ([ML] p.109).

Definition 18 A category is complete if it has all small limits and cartesian if it has all finite limits. Dually it is cocomplete if it has all small colimits and cocartesian if it has all finite colimits. A bicomplete category is both complete and cocomplete.

The relationship between (co)limits and adjoint functors can be expressed by the following :

Proposition 3 If $F:A \longrightarrow B$ is left adjoint to $G:B \longrightarrow A$, then G preserves limits and F preserves colimits.

Proof : ([LS] Prop.5.10 p.25).

Definition 19 A functor which preserves all finite limits is said to be left exact.

To clarify notation vis-à-vis (co)products :

Definition 20 Let $f:A \longrightarrow A'$ and $g:B \longrightarrow B'$ be arrows then a product of arrows f and g is $f \times g:A \times B \longrightarrow A' \times B'$, given by the rule $(a,b) \longmapsto (f(a),g(b))$.

Definition 21 Let $f:A \longrightarrow B$ and $g:A \longrightarrow C$ be arrows then the product arrow $\langle f,g \rangle:A \longrightarrow B \times C$ is given by the rule

$$a \longmapsto (f(a),g(a)).$$

Definition 22 Let $f:A \longrightarrow A'$ and $g:B \longrightarrow B'$ be arrows then the coproduct of arrows f and g is $f + g:A + B \longrightarrow A' + B'$ given by the rule $a \in A \longmapsto f(a)$ and $b \in B \longmapsto g(b)$.

Definition 23 Let $f:B \longrightarrow A$ and $g:C \longrightarrow A$ be arrows then the coproduct arrow $[f,g]:B + C \longrightarrow A$ is given by the rule

$$b \in B \longmapsto f(b) \text{ and } c \in C \longmapsto g(c).$$

Later we shall see that categories of 'spaces' need to be cartesian closed.

Definition 24 A cartesian closed category is a cartesian category \underline{C} (thus having products), such that for each object B of \underline{C} , the functor $(-) \times B: \underline{C} \rightarrow \underline{C}$ (sending an object X in \underline{C} to its product $X \times B$) has a right adjoint $(-)^B: \underline{C} \rightarrow \underline{C}$.

This adjunction means that for all objects A, B , and C in \underline{C} there is an isomorphism

$$\lambda: \underline{C}(A \times B, C) \xrightarrow{\sim} \underline{C}(A, C^B).$$

The latter is often given the mysterious name of ' λ -conversion'.

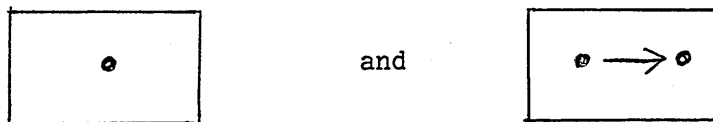
However, this only means that λ sends any map $g: A \times B \rightarrow C$ to its unique exponential adjoint $\hat{g}: A \rightarrow C^B$. In a cartesian closed category, the class of arrows $\underline{C}(A, B)$ enriches to an exponential object denoted by B^A in \underline{C} . If $ev: C^B \times B \rightarrow C$ is the counit of the adjunction, then we expect the following diagram to commute

$$\begin{array}{ccc} & C^B \times B & \\ \hat{g} \times id_B \uparrow & \searrow ev & \\ A \times B & \xrightarrow{g} & C \end{array}$$

If category \underline{C} is locally small then there is a functor $Hom: \underline{C}^{op} \times \underline{C} \rightarrow \underline{Set}$. The value of this functor for a pair of objects (A, B) in \underline{C} is $hom_{\underline{C}}(A, B)$ (or $\underline{C}(A, B)$). For a product of arrows $g \times h: (A, B) \rightarrow (A', B')$ (where $g: A' \rightarrow A$ and $h: B \rightarrow B'$ are arrows in \underline{C}) then $Hom(g, h)$ sends $f \in \underline{C}(A, B)$ to $hgf \in \underline{C}(A', B')$. If F and G are functors in a locally small category, we denote the set of natural transformations between them by $Nat(F, G)$.

Definition 25 If \underline{C} is locally small, a functor $F: \underline{C} \rightarrow \underline{Set}$ is said to be representable if F is isomorphic to $Hom_{\underline{C}}(C, -)$ for some

objects C in \underline{C} . Dually F is representable when isomorphic to $\text{Hom}_{\underline{C}}(-, C)$ and F is a functor from \underline{C}^{op} to Set . We shall denote $\text{Hom}_{\underline{C}}(C, -)$ and $\text{Hom}_{\underline{C}}(-, C)$ by h^C and h_C respectively. For example in Graph, the following functors (= graphs)



are representables, as they are isomorphic to h_U and h_I (see Definition 14, 25).

Of considerable importance is Yoneda's Lemma.

Proposition 4 If \underline{A} is locally small and $F: \underline{A}^{op} \rightarrow \text{Set}$ is a functor, then $\text{Nat}(h_A, F)$ is in one-to-one correspondence with $F(A)$ for all objects A in \underline{A} .

Proof : ([LS] Prop.2.7 pp.10-11).

Proposition 5 The Yoneda functor

$$Y: \underline{C} \rightarrow \text{Set}$$

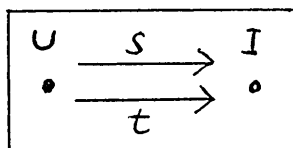
defined by

$$C \in \underline{C} \mapsto h_C \quad \text{and} \\ f: C' \rightarrow C \mapsto h_f: h_{C'} \rightarrow h_C$$

is a full embedding.

Proof : ([LS] Cor.2.9 p.11).

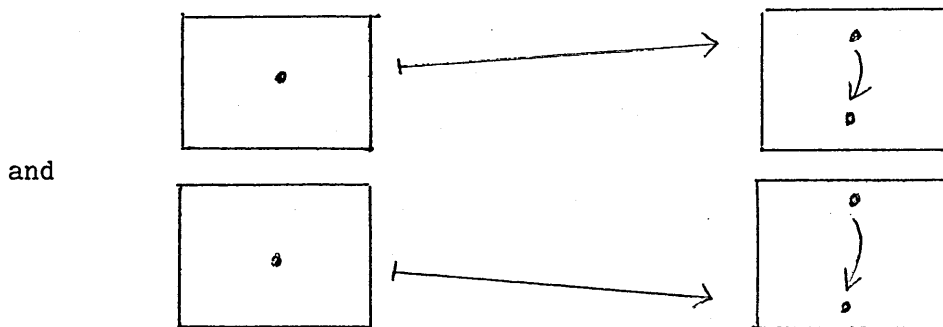
This allows us to assert that \underline{C} is equivalent to a full subcategory of $\text{Set}^{\underline{C}^{op}}$. For example, the Yoneda functor embeds



into Graph, by sending U to h_U and I to h_I . This can be pictured as follows :

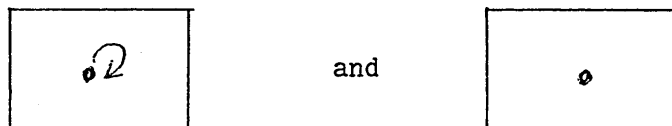


Also $s:U \longrightarrow I$ and $t:U \longrightarrow I$ are sent to



respectively. It will be seen later that the Yoneda functor permits a fruitful abuse in confusing \underline{C} with its 'copy' $Y(\underline{C})$ in $\underline{\text{Set}}^{\underline{C}^{op}}$.

It can readily be seen that the terminal object of Graph and RGraph can be pictured as



respectively. 1 is a representable in RGraph, but not in Graph.

However, the condition $1 \in \underline{C}$ may not for some \underline{C} be strictly true, yet may be true up to 'splitting idempotents'. There is for some categories \underline{C} a Cauchy completion $K(\underline{C})$, such that

$$\underline{C} \subseteq K(\underline{C}) \subseteq \underline{\text{Set}}^{\underline{C}^{op}} \cong \underline{\text{Set}}^{K(\underline{C})^{op}}.$$

To explore this we need the following.

Definition 26 Any arrow $f:A \longrightarrow A$ in a category \underline{C} is an idempotent if $f.f = f$. An idempotent is said to split when there exist arrows $g:A \longrightarrow B$ and $h:B \longrightarrow A$ such that $f = h.g$ and $g.h = \text{id}_B$.

Definition 27 For any category \underline{C} , the Cauchy completion or

Karoubi envelope (denoted by $K(\underline{C})$) is a category in which

- (i) objects are the idempotent arrows of \underline{C} ; and
- (ii) arrows $f \dashrightarrow g$ (where $ff=f$ and $gg=g$ in \underline{C}) are the triples

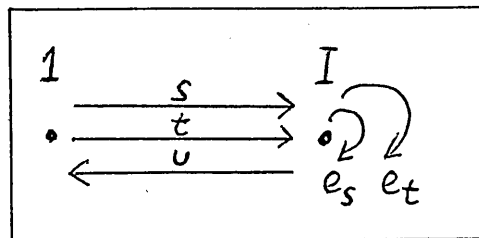
(f, ϕ, g) , where $\phi : A \dashrightarrow B$ is an arrow in \underline{C} such that $\phi \cdot f = \phi = g \cdot \phi$, or equivalently $g \cdot \phi \cdot f = \phi$.

Proposition 6 For any category \underline{C} , the Karoubi envelope $K(\underline{C})$ is a category in which all idempotents split.

Proof : ([BU] Prop.3.2).

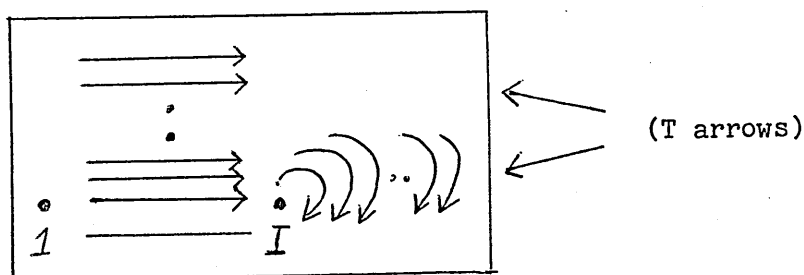
We shall only be interested in those Cauchy completions of small categories \underline{C} , such that presheaves on \underline{C} and $K(\underline{C})$ are equivalent.

For example, the \underline{M} -set, equivalent to \underline{RGraph} , described under Definition 14 is a category in which 1 is not representable. By obtaining the Karoubi envelope of the monoid M (considered as a category) a more plastic version of the category of \underline{M} -sets is obtained in which 1 is representable. Now it can readily be seen from the multiplication table that $st = s$, $ss = s$, $ts = t$, and $tt = t$ so that the monoid M is actually a band in which for all elements (other than the unit) $e_i \cdot e_j = e_i$ ($i, j = 1, 2$). Next treat the multiplication table of the monoid M as composition of arrows in a category \underline{M} and form the Karoubi envelope $K(\underline{M})$. Simple calculations show that $K(\underline{M})$ has 3 objects and 13 arrows. However, two of the objects are isomorphic and there are only seven non-equivalent arrows. This results in the category $\underline{M}(2)$ equivalent to $K(\underline{M})$ which we picture as



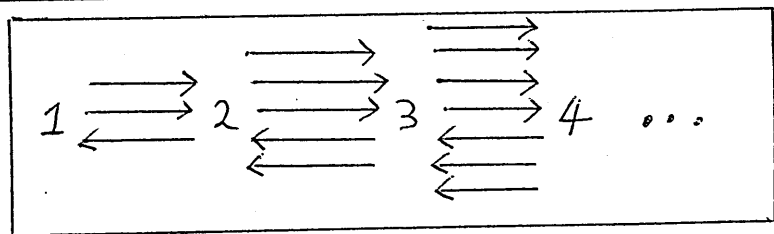
in which 1 is representable. The category $\underline{Set}^{\underline{M}(2)^{cp}}$ is equivalent

to RGraph. Elements of a variable set at stage 1 are nodes, and elements at stage I are edges. The restriction maps s, t describe the source and target of each edge, whereas the restriction map $u: I \rightarrow 1$ takes nodes to their associated degenerate loops. The monoid action on the set of edges is recovered through the variation carried along e_s and e_t . In principle, we can generalise graphs by taking any band in which $e_i \cdot e_j = e_i$ for $i, j = 1, \dots, T$. By taking the Karoubi envelope of these bands, we obtain categories $\underline{M}(T)$, which are pictured like $\underline{M}(2)$



but with T arrows instead of 2. Presheaves on $\underline{M}(T)$ may be usefully thought of as generalised reflexive graphs.

The same technique of Cauchy completion can be applied to the category whose objects consist of the natural numbers $(1, 2, 3, \dots)$ and whose morphisms consist of the (monoid of) monotone endomaps. The Karoubi envelope is a category $\underline{\Delta}$, known as the simplicial category and pictured as



in which the monoid actions on objects have been suppressed (see [ML] pp.172-3). The variable sets of Set $\underline{\Delta}^{c.p}$ are known as simplicial sets. Algebraic topologists often use these simplicial sets to model spaces by affine simplices. They rewrite the objects $\{1, 2, 3, 4, \dots\}$ as $\{0, 1, 2, 3, \dots\}$ using the geometric dimension ([ML]

p.174). Truncation at dimension 2 results in variable sets in $\text{Set}^{\Delta^{op}}_2$ which can be used to model triangulated surfaces ([LWQ] pp.26-7). It can easily be shown that the one-dimensional simplicial sets in $\text{Set}^{\Delta^{cp}}_1$ are equivalent to reflexive graphs in $\text{Set}^{M(2)^{cp}} \simeq \text{RGraph}$ ([LWQ]).

The variable sets X in $\text{Set}^{\underline{C}^{cp}}$ where \underline{C} is a small category can also be characterised as a special class of small categories known as fibrations. Since \underline{C} is small, its structure has a basic description in terms of sets of objects and arrows. Given \underline{C} , let C_o be the set of objects and C_1 the set of arrows of \underline{C} . Also we have set functions $\text{dom}: C_1 \rightarrow C_o$ and $\text{cod}: C_1 \rightarrow C_o$ indexing the domain and codomain of each arrow in \underline{C} . Now X in $\text{Set}^{\underline{C}^{cp}}$ may well be described by a right action mediated by the arrows of \underline{C} . As a variable set, X can be parameterised by the objects of \underline{C} , by rewriting the elements of X as a function $\phi: X_o \rightarrow C_o$ which indexes the stage of definition of the elements of X . We call X_o the set of objects of a category \underline{X} . The set of arrows X_1 can be formed by the following pullback (in Set)

$$\begin{array}{ccc} & \rho_2 & \\ X_1 & \xrightarrow{\quad} & C_1 \\ \rho_1 \downarrow & & \downarrow \text{cod} \\ X_o & \xrightarrow[\phi]{} & C_o \end{array} .$$

So that $X_1 = X_o \times_{C_o} C_1 = \{(x, \lambda) \mid \phi(x) = \text{cod}(\lambda)\}$, where x is an element of X at stage C and $\lambda: C' \rightarrow C$ is an arrow in \underline{C} . The right action of (x, λ) on x sends x to an element x' defined at stage C' , subject to the following conditions

- (i) $x.(x.\text{id}_C) = x$,
- (ii) $x.(x, \lambda).(x', \mu) = x.(x, \lambda.\mu)$ whenever $C'' \xrightarrow{\mu} C' \xrightarrow{\lambda} C$

in \underline{C} .

To be sure, the projection $p_1 : X_1 \longrightarrow X_0$ onto the first factor describes the codomain of the arrows in X_1 , and the right action indexes the domain.

Definition 28 The pairs of set (X_0, X_1) where X_1 is formed by pulling back $\phi : X_0 \longrightarrow C_0$ and $\text{cod} : C_1 \longrightarrow C_0$ (where C_0 and C_1 are the sets of objects and arrows of \underline{C}) constitute a full subcategory of the slice category Cat/\underline{C} known as the discrete fibrations, and denoted by $\text{Dfib}(\underline{C})$. As an object in Cat/\underline{C} the pair (X_0, X_1) form a category denoted by \underline{X} .

Proposition 7 Variable sets known as presheaves are equivalent to the discrete fibrations; that is,

$$\text{Set}^{\underline{C}^{\text{op}}} \cong \text{Dfib}(\underline{C}) \subset \text{Cat}/\underline{C} \subset \text{Cat}.$$

Proof : See ([BW] pp.228-9) where the proof is given for the equivalence between $\text{Set}^{\underline{C}}$ and the discrete opfibrations. The reader will have no trouble in rewriting the proof for $\text{Set}^{\underline{C}^{\text{op}}}$ and fibrations.

Proposition 8 The slice category $(\text{Set}^{\underline{C}^{\text{op}}})/\underline{X}$ where \underline{X} is a presheaf is equivalent to $\text{Set}^{\underline{X}^{\text{op}}}$, where \underline{X} is the discrete fibration associated with \underline{X} in Cat/\underline{C} .

Proof : ([JT1] p.50).

The dialectical twist here is that variable sets are not just a generalisation of the idea of sets, they are also categories.

CHAPTER SIX

TYPE THEORY AS THE FORMAL

"... logic in its traditional form is a purely formal science, and thus in any specific use made of it in the sciences or elsewhere, it is one and the same; the life which it assumes for the knower in such use is its proper life."

H-G Gadamer

In Chapter One, I asserted that the 'mathematical subsystem' could be disaggregated into the Formal and the Conceptual with an act of interpretation between them (Fig.2). I have two objectives in this Chapter. Firstly, to outline type theory as the most powerful version of the Formal for Science. Secondly, to show (through a simple example of the application of elementary differential calculus to Galileo's studies of falling bodies at the dawn of modern science) that the type theories needed by Science may be non-classical, meaning that models will have to be found in set-like categories markedly different from Set.

I shall start with Lawvere's world-picture of mathematical activity.

Lawvere's World Picture

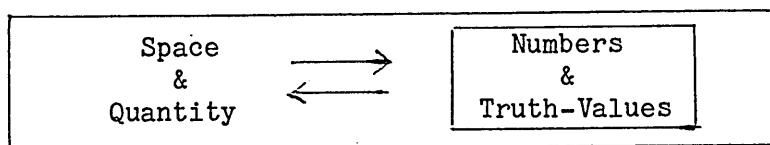


Figure 5

He has argued that ([LWP] p.378) :

"... the essential object of study in mathematics is space and quantitative relationships. Thus, as an essential part of the scientific world-picture, we have the mathematical world-picture (see Fig.5) whose links with the remainder of the scientific world-picture should never be forgotten"

In this picture, our experience of the objects in 'the external world' is mediated through the concepts of space and quantity.

These concepts are operationalised in terms of numbers and truth values, and at the heart of mathematical activity is the

elucidation and explication of mathematical concepts. Lawvere claims ([LWA] p.281) :

"That pursuit of exact knowledge which we call mathematics seems to involve in an essential way two dual aspects, which we may call the Formal and the Conceptual. For example, we manipulate algebraically a polynomial equation and visualize geometrically the corresponding curve. Or we may concentrate in one moment on the deduction of theorems from the axioms of group theory, and in the next consider the actual classes of groups to which the theorems refer. Thus the Conceptual is in a certain sense the subject matter of the Formal".

Here the mathematician visualises or conceives of entities such as spaces, graphs, bundles, groups, rings, modules, and categories. He uses axioms or sentences in some formal language to gain conceptual control over these notions; essentially the concepts are to be exhibited as models of a theory. The necessity of 'formality' is usually expressed in such terms as 'precision', 'lucidity', and 'clarity', and the passage between the Formal and the Conceptual relates what can be coherently visualised as a concept to what can be sensibly said about it.

Now mathematical knowledge is in no respect different to other forms of knowledge in that it is mediated through language. Searle has argued that knowledge, to count as knowledge, must be clearly expressible. Assuming a speaker (e.g. a mathematician) expresses his intention precisely, then it is possible (in

principle) for every speech-act carried out to be specified by a complex sentence (in ordinary language). Searle's Principle of Expressibility can be put formally in the following way ([SE] p.20):

For every meaning Z, it is the case that, if there is a speaker X in a language community C, then it is possible that there is an expression E in the language spoken by C which is an exact expression of Z.

When knowledge is at issue that possibility is convertible to necessity. This Principle can help us to see that the passage from the Formal to the Conceptual is one of finding an exact expression for a concept. Formal languages are not just syntactical collections of symbols to be manipulated for their own sake (as in Formal Logic), but precise tools for the utterance of speech-acts with mathematical and scientific contents.

In the following exposition of type theory, I shall assume the reader is acquainted with first-order logic in which theories are written down as axioms and structured sets are produced as models of these axioms. The axioms usually use the logical symbols of the classical propositional calculus on the set {true,false} (denoted by 2), together with the quantifiers \forall ('for all'), \exists ('for some'), and $\exists!$ ('for some unique'), various brackets for clarity, and non-logical symbols which act as predicates or relations. Type theories are just an extension to this general idea. Axioms remain. However, as many types as will be needed are added to the single type of first-order logic. The need for type theory arises from the distinctions we make in ordinary language. As Lambek and Scott point out ([LS] p.125) :

"Types are inherent in everyday language, for example, when we distinguish between 'who' and 'what' or between 'somebody' and 'something' ".

When we make these distinctions, we may, for example, postulate types A and B, and if t is a term of type A and s is a term of type B, then we may by abuse say $t \in A$ and $s \in B$. By making these distinctions in terms of types, we already have a useful extension to first-order logic. However, type theory is not like first-order logic in that it allows higher-order types. As Freyd puts it ([FR] p.6) :

"... First-order logic is surely an artifice, albeit one of the most important inventions in human thought. But none of us thinks in first-order language. The predicates of natural dialectics are order-insensitive (one moment's individuals are another's equivalence-class) and our appreciation of mathematics depends on our ability to interpret the words of mathematics. The interpretation itself is not first-order."

The value of type theory is that it permits higher-order types, which fit more closely the natural dialectics of language.

The expressive power of formal systems described in type theory is great. After a little practice, we can easily acquire the feeling that all mathematics can be formalised in them. Type theory is the most powerful representation of the Formal available to us, for it offers us both discrimination between types (which is essential for Science) and the possibility of universal and existential quantification (which is essential for the law-like hypotheses of Science). Accordingly, I shall think of the formalised fragments of scientific discourse (making demands on Mathematics) as statements in a type theory requiring interpretation in a set-like category.

As the basis for my outline of type theory, I follow closely the formal system of Lambek and Scott ([LS]). However, my exposition differs in two main respects. Firstly, 'a seemingly

stronger version' of type theory needs function types ([LS] p.132), and since even the most elementary examples from Science require function types I add in various fragments of a type theory with function types as suggested by de Vries ([VR] Chap.1). Secondly, I exclude all the usual axioms about Peano arithmetic and the natural numbers from my discussion of type theory. It is not that I think these axioms are not a part of type theory, but reasons of space and the need to concentrate on an exposition of type theory useful to Science lead me to suppress these features of elementary arithmetic, which we can tacitly regard as there. Furthermore, I shall make the usual assumptions about the conventions regarding free and bound variables and the renaming variables in substitution by terms, and will skate over the complexities of these problems ([JT4] pp.19-23). These complexities would only obscure the picture of type theory outlined here, and hinder the understanding of the working mathematician/scientist. The super-logical reader will readily incorporate the adjustments as a matter of course.

Intuitionist type theory can be described as follows.

Definition 1 The kernel of a type theory consists of the following data.

- a) A set of basic type symbols A, B, C, \dots including the special types 1 ('one') and Ω ('truth type').
- b) The smallest set of type symbols (including the basic types) closed under product and power-set formation :
 - i) if A and B are types, then so is $A \times B$;
 - ii) if A is a type, then so is $P(A)$;
 - iii) Ω will denote the type $P(1)$.
- c) A set of function symbols f, g, h, \dots where each function symbol has type symbols as domain and codomain, denoted by $f: A \rightarrow B$.

d) A set of axioms formulated in the terms and formulae of the language constructed from the above data.

Definition 2 The terms of a type theory are freely generated from the types as follows.

- a) $*$ is a term of type 1, and for each type A we have countably many variables x_1, x_2, x_3, \dots of type A. We shall denote 'let x be a variable of type A' by $x \in A$. Variables of type $P(A)$ will be denoted by upper-case letters e.g. $S \in P(A)$. We tacitly assume all the conventions regarding free and bound variables (see Definition 3).
- b) $\langle s, t \rangle$ is a term of type $A \times B$ for all terms s of type A and t of type B.
- c) $\{x \in A \mid \phi(x)\}$ is a term of type Ω , if $\phi(x)$ is a term of type Ω and ϕ is a predicate that can be applied to variables of type A.
- d) $t \in S$ is a term of type Ω for each term t of type A and S of type $P(A)$.
- e) $f(t)$ is a term of type B for each term t of type A and function symbol $f: A \rightarrow B$.
- f) \top ('true') and \perp ('false') are terms of type Ω .
- g) If p and q are terms of type Ω , then so are $p \wedge q$ ('p and q'), $p \vee q$ ('p or q'), and $p \Rightarrow q$ ('if p then q').
- h) If $\phi(x)$ is a term of type Ω , then $(\forall x \in A) \phi(x)$ and $(\exists x \in A) \phi(x)$ are terms of type Ω . The symbols \forall ('for all') and \exists ('for some') are quantifiers.

Definition 3 Terms of type Ω are called formulae. The variable x in $\phi(x)$ is said to be bound in quantified formulae and terms like $\{x \in A \mid \phi(x)\}$. A variable is said to be free when it is not bound.

Definition 4 A full type theory is obtained by expanding the kernel with further symbols - \neg ('not'), \Leftrightarrow ('equivalence'), $=$ ('equals'), $\{x\}$ ('singleton'), $\exists!$ ('for some unique'), \subseteq ('inclusion'). The symbol $\exists!$ is also a quantifier. We define the extra symbols in terms of the kernel.

- a) $\neg p$ is $p \Rightarrow \perp$.
- b) $p \Leftrightarrow q$ is $(p \Rightarrow q) \wedge (q \Rightarrow p)$.
- c) $a = a'$ is $(\forall S \in P(A))[a \in S \Leftrightarrow a' \in S]$.
- d) $\{a\}$ is $\{x \in A \mid a = x\}$.
- e) $(\exists! x \in A) \phi(x)$ is $(\forall x' \in A)\{x \in A \mid \phi(x)\} = \{x'\}$.
- f) $S \subseteq T$ is $(\forall x \in A)(x \in S \Rightarrow x \in T)$

where S and T are terms of type $P(A)$.

The axioms and formulae of a type theory are governed by a relation of entailment, denoted by $\vdash\text{---}X$, between terms of type Ω . Entailment is defined for each finite set X of variables between terms of type Ω , whose free variables are contained in X (see [LS] pp.130-1). We adopt the conventions $\vdash\text{---}$ for $\vdash\text{---}\emptyset$ and $\vdash\text{---} p$ for $\top \vdash\text{---} p$.

Definition 5 The set of axioms and rules of a type theory contains the axioms of the data defining the specific type theory, together with the following standard set of axioms and rules.

a) Structural Rules

- 1) $p \vdash\text{---}X p$.
- 2)
$$\frac{p \vdash\text{---}X q \quad q \vdash\text{---}X r}{p \vdash\text{---}X r}.$$
- 3)
$$\frac{p \vdash\text{---}X q}{p \vdash\text{---}X \bigcup \{y\} q}.$$

$$\sim 4) \frac{\phi(y) \vdash\text{---} X \cup \{y\} \psi(y)}{\phi(b) \vdash\text{---} X \psi b}$$

if y is a variable of type A and b is a term of type B , with the usual tacit conventions about free and bound variables.

b) Logical Rules

$$5) i) p \vdash\text{---} X T ;$$

$$ii) \perp \vdash\text{---} X p.$$

$$6) i) r \vdash\text{---} X p \wedge q,$$

$$\text{iff } r \vdash\text{---} X p \text{ and } r \vdash\text{---} X q.$$

$$ii) p \vee q \vdash\text{---} X r,$$

$$\text{iff } p \vdash\text{---} X r \text{ and } q \vdash\text{---} X r.$$

$$7) p \vdash\text{---} X q \Rightarrow r \text{ iff } p \wedge q \vdash\text{---} X r.$$

$$8) p \vdash\text{---} X (\forall y \in B) \psi(y)$$

$$\text{iff } p \vdash\text{---} X \cup \{y\} \psi(y).$$

$$9) (\exists y \in B) \psi(y) \vdash\text{---} X p$$

$$\text{iff } \psi(y) \vdash\text{---} X \cup \{y\} p.$$

c) Comprehension

$$10) \vdash\text{---} X (\forall x \in A)[x \in \{x \in A \mid \phi(x)\} \Leftrightarrow \phi(x)].$$

d) Extensionality

$$11) \vdash\text{---} (\forall S, T \in P(A))(\forall x \in A)[(x \in S \Leftrightarrow x \in T) \Rightarrow S = T].$$

$$12) \vdash\text{---} (\forall s, t \in \Omega)[(s \Leftrightarrow t) \Rightarrow s = t].$$

There are other rules relating entailment to products ([LS]

pp.131), but the details will be omitted here.

Proposition 1 The following are consequences of the foregoing rules and axioms.

1) The usual rules of intuitionist propositional logic :

$$a) p \Rightarrow (q \Rightarrow p).$$

$$b) [p \Rightarrow (q \Rightarrow r)] \Rightarrow [(p \Rightarrow q) \Rightarrow (p \Rightarrow r)].$$

$$c) (p \wedge q) \Rightarrow p; (p \wedge q) \Rightarrow q.$$

d) $p \Rightarrow [q \Rightarrow (p \wedge q)]$.

2) Axioms for symmetry and transitivity of = .

e) $(\forall x, y \in A)(x = y \Rightarrow y = x)$.

f) $(\forall x, y, z \in A)(x = y \wedge y = z \Rightarrow x = z)$.

3) Three rules of inference.

g)
$$\frac{p, \quad p \Rightarrow q}{q}$$

provided that the free variables of p appear free in q .

h)
$$\frac{p \Rightarrow \phi(x)}{p \Rightarrow (\forall x \in A) \phi(x)}$$

provided x does not appear free in p .

j)
$$\frac{\phi(x) \Rightarrow q}{(\exists x \in A) \phi(x) \Rightarrow q}$$

provided x does not appear free in q .

4) Two additional axioms for quantifiers.

k) $(\forall x \in A) \phi(x) \Rightarrow \phi(x) = T$.

l) $\phi(x) \wedge (\exists x \in A) T \Rightarrow (\exists x \in A) \phi(x)$,

where $(\exists x \in A) T$ indicates that type A is inhabited or non-empty.

Proof : ([VR] pp.5-6, [LS] pp.128-39).

The full type theory outlined in the previous definitions is intuitionist type theory. It is well-known that among the tautologies of intuitionist logic the following propositions, $p \vee \neg p$ and $\neg \neg p \Rightarrow p$, are not provable, and this applies to intuitionist type theory too.

On the other hand, $p \vee \neg p$ and $\neg \neg p \Leftrightarrow p$ (or in view of Definition 4, $(\neg \neg p \Rightarrow p) \wedge (p \Rightarrow \neg \neg p)$) are tautologies in classical propositional logic. Since $\neg \neg p \Rightarrow p$ cannot be

established in intuitionist logic, it follows that $\neg\neg p \Leftrightarrow p$ is not a theorem of intuitionist logic. We use this to distinguish intuitionist and classical logic with a further axiom.

Definition 6 A type theory is said to be classical when

$$(\forall t \in \Omega)(t \vee \neg t) \text{ or equivalently } (\forall t \in \Omega)(\neg\neg t \Rightarrow t).$$

A formulation of type theory without explicit function types is somewhat unsatisfactory for the needs of Science. It would be desirable to have available a notion of a function, such that whenever we encounter a predicate $\phi(x,y)$ we can prove that it represents a function through the statement

$$(\forall x \in A)(\exists ! y \in B) \phi(x,y).$$

To incorporate an explicit type for functions, we can extend the language in the following way ([VR] Chap.1).

Definition 7 For a type theory with function types.

- a) Add to the rules, that define the formation of types, the new rule : closure under exponentials. That is, for a function symbol $f:A \rightarrow B$, define the exponential type B^A , and regard f as a term of type B^A .
- b) Add to the rules, that determine terms, the rule of closure under application. That is, if f is a term of type B^A , and if x is a term of type A , then $f(x)$ is a term of type B . A term of type B^A can be denoted by $(\lambda x \in A) \phi(x)$, which refers to the rule $x \mapsto \phi(x)$ (where $\phi(x)$ is a term of type B) determining the function $f:A \rightarrow B$.
- c) Add the following axiom concerning extensionality of functions and the axiom formulating the unique choice of functions.
 - i) $(\forall f,g \in B^A)[f = g \Leftrightarrow (\forall x \in A)f(x) = g(x)]$.
 - ii) $[(\forall x \in A)(\exists ! y \in B) \phi(x,y)] \Rightarrow (\exists ! f \in B^A)(\forall x \in A) \phi(x,f(x)).$

As an example of a type theory, consider the following canonical type theory of the category Set.

Definition 8 The internal language of Set, denoted by $L(\text{Set})$, is a classical type theory and has as basic type symbols the sets A, B, C, \dots , and function symbols are the set functions f, g, h, \dots . In particular, the type 1 is the singleton set $\{0\}$, and type Ω is the set $\{\text{true}, \text{false}\}$ (denoted by 2) endowed with the structure of classical propositional calculus. \top is true and \perp is false. Physicists and engineers often treat a variable x (in calculus) as a variable quantity. This approach is good enough for type theory and logic too. Variables can be treated as indeterminates ([LA]) and to every possible predicate (or polynomial) $\phi(x)$ (in the variable x of type A with values in B) we can associate a function $f: A \rightarrow B$ (where B is usually the set 2). Furthermore, since Set is cartesian closed, to every $f: A \rightarrow B$ we can associate an exponential adjoint

$$[f]: 1 \rightarrow B^A$$

called the name of f , such that the following diagram commutes :

$$\begin{array}{ccc}
 & B^A \times A & \\
 & \uparrow & \searrow \text{ev} \\
 [f] \times \text{id}_A & 1 \times A & \xrightarrow{f \cdot p_2} B
 \end{array}$$

We write $f(x) = \phi(x)$ and call this functional completeness.

Now $L(\text{Set})$ has as terms of type A in the variables x_i of type A_i ($i = 1, \dots, n$) predicates $\phi(x_1, \dots, x_n)$. In particular

- a) $*$ is a variable of type 1;
- b) $\langle a, b \rangle$ is a variable of type $A \times B$;
- c) S is a variable of type $P(A)$.

Furthermore, we can associate any predicate

d) $a = a'$ of type Ω with a function

$$\begin{aligned} \delta_A : A \times A &\longrightarrow 2, \\ \langle a, a' \rangle &\longmapsto \begin{cases} \text{true if } a = a' \\ \text{false otherwise;} \end{cases} \end{aligned}$$

e) $a \in A$ of type Ω with a function

$$\begin{aligned} \text{ev} : P(A) \times A &\longrightarrow 2, \\ \langle S, a \rangle &\longmapsto \begin{cases} \text{true if } a \in S \\ \text{false otherwise;} \end{cases} \end{aligned}$$

f) $\{x \in A \mid \phi(x)\}$ of type $P(A)$ with the function

$$\overline{f} : 1 \longrightarrow 2^A \text{ which picks out that function}$$

$$(\lambda x \in A) \phi(x) \text{ which defines}$$

$$f : A \longrightarrow 2, \text{ that is}$$

$$x \longmapsto \begin{cases} \text{true if } \phi(x) \\ \text{false otherwise.} \end{cases}$$

(More generally for any predicate $\phi(x)$ in the variable x of

type A with values in B is associated a function $f : A \longrightarrow B$

with exponential adjoint given by the name of f .)

Although $P(A)$ and 2^A are different types, it is implicit that they are equivalent. It is also easy to see that the type C^A is equivalent with the type of composable functions

$$\{\langle g, f \rangle \in C^B \times B^A \mid g \circ f = h \in C^A\}.$$

How are we to interpret the entailment relation in $L(\underline{\text{Set}})$?

Definition 9 Let X be a set of variables. Let $\text{true}_C : C \longrightarrow 2$ be the unique arrows from all sets C to 2 such that $\text{true}_C(c) = \text{true}$ for all $c \in C$. Let

$$\phi_1(x), \phi_2(x), \dots, \phi_n(x) \vdash\!\!\!\vdash x \phi_{n+1}(x)$$

be an entailment between terms of type Ω , and let $f_i : A \longrightarrow 2$

be the unique arrows $(\lambda x \in A) \phi_i(x)$ ($i = 1, \dots, n+1$), then

entailment means that for all arrows $h : C \longrightarrow A$ and for all sets C

where $f_i . h = \text{true}_C$, then $f_{n+1} . h = \text{true}_C$.

For example, let X be $\{x\}$, and let $f:A \rightarrow 2$ and $g:A \rightarrow 2$ be the arrows given by $(\lambda x \in A) \phi(x)$ and $(\lambda x \in A) \psi(x)$, then

$\phi(x) \vdash\text{---} x \psi(x)$ means that $f.h = \text{true}_C$ implies $g.h = \text{true}_C$ for all functions $h:C \rightarrow A$. It can easily be seen that $\{x \in A | f(x) = \text{true}\}$ is a subset of $\{x \in A | g(x) = \text{true}\}$.

Definition 10 $\vdash\text{---} x \phi(x)$ means ϕ holds in $L(\text{Set})$, and we write $\text{Set} \models \phi(x)$ and $\phi(x) = \top$.

By axiom 8 in Definition 5,

$$\vdash\text{---} x \phi(x) \text{ means } \vdash\text{---} (\forall x \in A) \phi(x),$$

so there is a unique function $f_\phi:A \rightarrow 2$, such that $f_\phi(x) = \text{true}$ for all a in A .

The internal language of Set can be put to work to prove familiar propositions about set functions. For example :

Proposition 2 A function $f:A \rightarrow B$ is

a) a monomorphism (or injection) iff

$$\text{Set} \models (\forall x, x' \in A)(f(x) = f(x') \Rightarrow x = x');$$

b) an epimorphism (or surjection) iff

$$\text{Set} \models (\forall y \in B)(\exists x \in A)(f(x) = y), \text{ and}$$

c) an isomorphism (or bijection) iff

$$\text{Set} \models (\forall y \in B)(\exists ! x \in A)(f(x) = y).$$

Proof : ([LS] p.153).

When we put the internal language to work in this way, we see that it is just the ordinary language in which we talk about mathematics. Now the trained mathematician tries to achieve a mental condition in which the concepts of sets and functions and the formal language of $L(\text{Set})$ are so intertwined that he might almost forget the difference between them ([AT3] p.1).

Nevertheless, $L(\text{Set})$ and Set are different objects.

In Definition 8 I showed how to transfer the category Set into a type theory called $L(\underline{\text{Set}})$. I now show the rudiments of interpreting the axioms of a specific type theory into Set.

Definition 11 An interpretation of a type theory H into Set is a mapping $[] : H \dashrightarrow \underline{\text{Set}}$ taking types into sets such that,

$$[1] = \{0\}$$

$$[\Omega] = 2$$

$$[A \times B] = [A] \times [B]$$

$$[P(A)] = 2^{[A]}$$

$$[B^A] = [B]^{[A]} = \{f | f:A \dashrightarrow B\},$$

and a function $f:A \dashrightarrow B$ is mapped into a morphism

$$[f] : [A] \dashrightarrow [B] \text{ in } [B^A].$$

Axioms of the type theory H look like formulae in $L(\underline{\text{Set}})$, and there is an implicit assumption that the axioms of the type theory H are simply being added to $L(\underline{\text{Set}})$. Predicates and formulae have an interpretation as subsets and functions with codomain 2. I will not give details here, as they can be found in any standard text book on set theory and logic (see for example [JT4] pp.20-1).

For example, let H be ThC , the theory of small categories. It has two types A ('arrows') and O ('objects'), and three function symbols

i) $d:A \dashrightarrow O$ ('domain'),

ii) $c:A \dashrightarrow O$ ('codomain'),

iii) $i:O \dashrightarrow A$ ('identity'),

and a predicate K ('composition') on $A \times A \times A$.

Category theory has six axioms which formalise ordinary language statements.

$$1) (\forall f \in A)[d(i(c(f)))=c(f) \wedge c(i(d(f)))=d(f)].$$

'The domain of the codomain of f is the codomain of f , and the codomain of the domain of f is the domain of f '.

$$2) (\forall f, g, h, l \in A)[K(h, g, f) \wedge K(l, g, f) \Rightarrow h=l].$$

'Composition of arrows is unique when it is defined'.

$$3) (\forall f, g \in A)(\exists h \in A)[K(h, g, f) \Leftrightarrow c(f)=d(g)].$$

'The composition of g with f is defined iff the codomain of f is the domain of g'.

$$4) (\forall f, g, h \in A)[K(h, g, f) \Rightarrow d(h)=d(f) \wedge c(h)=c(g)].$$

'If h is the composition of g with f, then the domain of h is the domain of f and the codomain of g is the codomain of h'.

$$5) (\forall h \in A)[K(h, h, i(d(h))) \wedge K(h, i(c(h)), h)].$$

'For any h, the domain of h is a left-identity for h under composition, and the codomain is a right-identity'.

$$6) (\forall h, g, f, l, j, u, v \in A)[K(h, g, f) \wedge K(l, j, g) \wedge K(u, l, f) \wedge K(v, j, h) \Rightarrow u = v].$$

'Composition is associative when defined'.

An interpretation of the type theory ThC into Set sends the types A and 0 into two sets A and 0 which define the sets of arrows and objects of small category. The composition of the arrows as described by the six axioms means that there is a subset K contained in $A \times A \times A$ such that $K = \{(h, g, f) | h=g.f\}$ satisfying the axioms. The three function symbols have a simple interpretation as the domain, codomain, and identity mappings between the arrows and objects. Thus an interpretation (or model) of ThC is a small category (an object of Cat). For any formula p provable in ThC, we write $\text{ThC} \vdash p$. The following propositions are regarded as essential by logicians if the type theory is to be used consistently.

Proposition 3 (Soundness Theorem) Let H be a type theory.

If $H \vdash p$ then Set $\models p$.

Proposition 4 (Completeness Theorem) Let H be a type theory,

then $H \vdash p$ iff Set $\models p$.

Proof : ([GB] p.186 & p.227, [JT4] p.25).

Thus any theorem provable in ThC is valid for any object of Cat. And a formula in ThC is only provable if it is valid about all objects in Cat.

The example of ThC illustrates a general approach in mathematics. Let H be the theory of lattices, groups, rings, posets, topological spaces, or metric spaces, then an interpretation in Set produces an object in a concrete category. Similarly, let H be the theory of Peano arithmetic or Dedekind real numbers, then an interpretation in Set produces the natural or (Dedekind) real numbers respectively. Thus by 'adding' the axioms of H to L(Set), the Formal and the Conceptual can be related through the mapping of interpretation. It permits the discourse of mathematics to be formulated as precise statements in a type theory, and the concepts of mathematics can be regarded as (structured) relationships on sets. For as Blass puts it ([BS] p.5)

"It is a remarkable empirical fact that mathematics can be based on set theory. More precisely, all mathematical objects can be coded as sets ... and all their crucial properties can be proved from the axioms of set theory".

This remark indicates the power of set theory in its guise as type theory. It would be unsurprising that scientists did not formalise their scientific theories (in type theory) given that sets can code (empirically) the objects of the universe.

For a Science with a technical cognitive interest, the categorial framework of space, time, and substance suggests the types of E (= a space), T (= time), and B (= a material body). To describe motions of a body through space and time, we need a type theory with functions like $m: B \times T \rightarrow E$ and a statement like

$$(\forall m \in E^{B \times T}) (\forall \langle b, t \rangle \in B \times T) (\exists ! x \in E) m(\langle b, t \rangle) = x.$$

Thus already at this elementary level we can grasp the importance of type theory for Science, with a formal statement where the variables are interpreted as 'points' of a coordinate framework coding motion. Proceeding naively, Galileo could model a falling body ideally by a single point (the type 1), and both space and time could be typed by the geometric line (the type R). His experiments produced empirical descriptions of motion as $m \in R^R$ described by $s = 16t^2$ ($(\lambda t \in R) 16t^2$) (see Chapter 2, (1)). Interpreting these descriptions in Set, we can use the Dedekind reals to model the type R, and for every instant (or point) in time there is a unique coordinate (or point) specifying the location in space of the falling body.

However, the decision to interpret even this simple example of a physical theory in Set is not altogether a happy one, for by the 18th Century difficulties in interpreting statements were beginning to be felt ([DH] pp.242-3) :

"The leading problem was the connection between 'fluents' and 'fluxions', what would today be called the instantaneous position and the instantaneous velocity of a moving body ... In the case of the falling stone the fluent is given by the formula $s = 16t^2$, ... As the stone falls its velocity increases steadily. How can we compute the velocity of the falling stone at some instant of time, say at $t = 1$? We could find the average velocity for a finite time by the elementary formula : velocity equals distance divided by time. Can we use this formula to find the instantaneous velocity ? In an infinitesimal increment of time the increment of distance would also be infinitesimal; their ratio, the average speed during the instant, should be the finite instantaneous velocity we seek. We let dt stand for the infinitesimal increment of time and ds for the corresponding increment of distance ... We want to find the ratio ds/dt , which is to be finite. To find the increment of distance from $t = 1$ to $t = 1 + dt$ we compute the position of the stone when $t = 1$, which is $16 \times 1^2 = 16$, and its position when $t = 1 + dt$, which is $16 \times (1 + dt)^2$. Using a little elementary algebra, we find that ds , the increment of distance ... is $32dt + 16dt^2$. Thus the ratio ds/dt ... is equal to $32 + 16dt$... Have we solved our problem ? Since the answer should be a finite quantity, we should like to drop the infinitesimal term, $16dt$, and get the answer 32 feet per second, for the instantaneous velocity. That is precisely what Bishop Berkeley will not let us do."

Berkeley argued that neglecting the infinitesimal term $16dt^2$ in calculations was unintelligible. If a quantity is neglected, however small, we can no longer claim to have the exact velocity, but only an approximation. Either $dt = 0$ or $dt \neq 0$. By appealing to the law of the excluded middle, Berkeley argued that if $dt \neq 0$ then $32 + 16dt$ is not the same as 32. On the other hand, if $dt = 0$ then $ds = 0$ and the ratio ds/dt is not 32 but a meaningless expression $0/0$ ([DH] p.244). Clearly Berkeley's objections rest on an assumption that the logic of these matters is classical.

At that time Berkeley's classical logic could not be answered. To avoid inconsistency, mathematicians under the leadership of Weierstrass developed a rigorous approach to Analysis through set theory and classical logic. They achieved this by abandoning any reference to infinitesimals and any attempt to compute velocity as a ratio. Instead they defined velocity as a limit approximated by ratios of finite increments. Let Δs and Δt be finite increments of distance and time respectively. Then $\Delta s/\Delta t$ is the quantity $32 + 16\Delta t$. By choosing Δt sufficiently small we can make $\Delta s/\Delta t$ approximate values as close as we like to 32. This approach removes reference to non-finite numbers. It avoids setting Δt to 0 in the ratio $\Delta s/\Delta t$. It also avoids the apparent logical traps exposed by Berkeley, and allows the use of classical logic by differential geometers. Nevertheless there is a price to pay. The intuitively clear and physically measurable concept of instantaneous velocity becomes subject to the subtle and abstruse definition of a limit. For the sake of consistency we are led to a definition (limit) harder to grasp than our original concept (instantaneous velocity) ([DH] p.245).

However, engineers and physicists have never ceased using infinitesimals. Indeed Robinson has made infinitesimals respectable again through non-standard analysis ([DH] pp.246-54). This involves classical reasoning in a non-standard model of set theory (see Chap.7, Def.11 ff.). However Lawvere feels that non-standard analysis is 'counter-intuitive' ([LWP] p.387). He argues that the essential issues of differential calculus can be dealt with in 'a more natural fashion' provided we abandon classical set theory and the appeal to the law of the excluded middle ([LWV] pp.104-5).

This 'natural' approach assumes that any curve is a straight line in the infinitely small. When engineers postulate that

$$\text{distance}(t + dt) = \text{distance}(t) + \text{velocity}(t).dt \quad ,$$

then in order to neglect the term $16dt^2$ in $32dt + 16dt^2$ they must assume that $dt^2 = 0$. Taking this seriously, we are led to define

the type of infinitesimals (as a subtype of R) $D = \{d \in R \mid d.d = 0\}$.

Rewriting the assumption as 'any curve restricted to D is a

straight line', Kock formulates an axiom basic to the differential

calculus ([KK2] p.3) : for any $g:D \rightarrow R$ there exists a unique

$b \in R$ such that $(\forall d \in D) g(d) = g(0) + b.d$. The assumption of

enough nilpotents leads to immediate rigorous proofs ([LWP] p.384)

of all the basic derivations of calculus (Chain Rule, Leibniz'

Rule, fundamental theorem of calculus etc.), provided the fluxion

(derivative) f' of a fluent f is defined as being the (unique by

the axiom) one characterised by Taylor's formula

$$(\forall d \in D) f(x + d) = f(x) + f'(x).d \quad .$$

To be sure, this axiom is incompatible with classical logic. On the

one hand, using the Dedekind reals in Set is incompatible with the

axiom since D would consist of 0 alone. On the other hand, by using

the law of the excluded middle we can postulate the existence of

$h \in D$ with $h \neq 0$. Now Schanuel has constructed a function which is incompatible with the use of the law of the excluded middle in this way (see Note 3, p.227). Essentially this is a modern argument of Berkeley's case against fluxions, but it loses its destructive power when the unconditional validity of this logical law is surrendered ([RY] p.72). To retain the insights of engineers we must assume the derivations of calculus are examples of non-classical reasoning. So we may keep the axiom and abandon the law of the excluded middle ([KK2] p.5). The route followed by Weierstrass was to develop the different type theory of non-intuitive rigorous Real Analysis in the context of classical reasoning of Set. The alternative is to interpret theories about space and motion in set-like categories which are non-classical.

CHAPTER SEVEN

KEY ASPECTS OF ELEMENTARY TOPOI

"... just good enough to be applicable not only to sheaf theory, algebraic spaces ... as originally envisaged by Grothendieck ... but also to Kripke semantics, abstract proof theory and ... independence results in set theory."

F.W Lawvere

For my purposes, the categories which are set-like are the elementary topoi. I have two objectives in this Chapter. Firstly, to give a basic outline of key aspects of elementary topoi, and to show how they can be used (like Set) to serve as a suitable universe of discourse in which to elaborate our models of a type theory. Thus a topos is a suitable category for encoding a description of the Conceptual. Secondly, to reveal the sense in which topoi can be regarded as generalised spaces, as a preliminary exploration of notions which need to be distinguished from those of a topos of spaces.

Topoi can be approached through axioms expressed in the general language of category theory. Of course ([LWV]) p.118):

"There can be no doubt that in mathematical practice both sets and their membership as well as mappings and their composition play basic roles. But in setting up a formal theory one should also try to get clear on which of these is primary and which is secondary in mathematical practice."

In my view, composition of maps takes the leading role, since set theory poses considerable problems for the working mathematician.

As Lawvere puts it ([LWV] p.118-9) :

"The traditional view that membership is primary leads to a mysterious absolute distinction between x and $\{x\}$, to agonizing over whether or not the rational numbers are literally contained in the real numbers, to the 'discovery' that an ordered pair of elements in turn has elements which are, however, not the original elements, and to debates over whether the members of the natural number 5 are 0,1,2,3,4 or not, and all that is clearly just getting started; on its own formal face, a membership-based theory of sets is potentially littered with an infinite number of such formulas that even set theorists refrain from writing down due to their good mathematical sense. This situation, along with a very analogous situation with respect to the standard formalization of predicate logic, has led to the widespread view that a formalized theory and the calculations it tries to unify are necessarily so sharply divorced from each other that only a pedant would attempt to actually use a formalized set theory, which view only helps to isolate from most people the actual advances set theorists and logicians have made."

Thus the main problems of set theory would appear to be that it treats set-membership in vacuo as global and absolute. However, in practice mathematicians only usually consider set-membership as a relation between the elements of a given set and subsets of the same set; that is, set-membership is local and relative. In the latter context, membership can be described in terms of arrows $C \longrightarrow A \subseteq X$, where C parameterises some element and A is a subset of X . Thus to escape the paradoxes of set-theory, the local and relative notion of membership can be reduced to arrows and composition of maps. All this suggests that a purified approach to axiomatising the category Set in the language of category theory ([LWS]) may be more relevant to mathematical practice than the traditional approach through the formal face of set theory.

Axiomatisations of the category of sets are described by Lawvere ([LWS]) and Tierney ([TY]). The following axioms of topos theory are a subset of an elementary theory of the category of sets. This outline is based on Tierney ([TY]).

Definition 1 An elementary topos is any category \underline{E} , which satisfies the following three axioms.

E1) \underline{E} is cartesian and cocartesian.

E2) \underline{E} is cartesian closed.

E3) Subobjects in \underline{E} are representable; meaning that \underline{E} has a truth(-value) object Ω together with a subobject classifier $\text{true}: 1 \rightarrow \Omega$, such that for any monic $m: A \rightarrow X$, there is a unique characteristic map (or character) $\text{char}(m): X \rightarrow \Omega$, such that

$$\begin{array}{ccc} A & \xrightarrow{m} & X \\ \downarrow !_A & & \downarrow \text{char}(m) \\ 1 & \xrightarrow{\text{true}} & \Omega \end{array}$$

is a pullback in \underline{E} .

Before proceeding further, we need the following.

Definition 2 In a topos a monic m will be called the kernel of $\text{char}(m)$. Indeed any map $h: X \rightarrow \Omega$ will be the character of some monic which is its kernel. The kernel of h will be denoted by $\text{ker}(h)$. While characteristic maps are unique, kernels are unique only upto isomorphism. Thus

$$\text{char}(\text{ker}(h)) = h, \text{ and}$$

$$\text{ker}(\text{char}(m)) \cong m.$$

A map $\text{true}_A: A \rightarrow \Omega$, will be defined as the composition of $\text{true} \circ !_A$.

Definition 3 A set G of objects in a category \underline{C} is called a generating set if, for any two arrows $f: A \rightarrow B$ and $g: A \rightarrow B$, $f = g$ iff for all objects C in G and all $h: C \rightarrow A$, $fh = gh$.

Proposition 1 The representable functors h_C with C ranging over the objects of \underline{C} form a generating set for $\underline{\text{Set}}^{\underline{C}^{op}}$.

Proof : ([LS] Prop.9.2 p.171).

It is easily seen that for $\underline{\text{Set}} \cong \underline{\text{Set}}^1$ then $\{1\}$ forms a generating set.

Proposition 2 $\underline{\text{Set}}$ is a topos.

Proof : Tierney ([TY]).

To satisfy E1 it is enough that \underline{E} has an initial object, terminal object, equalisers, pullbacks, coequalisers, and pushouts.

Certainly it is well-known that all finite limits and colimits exist in $\underline{\text{Set}}$. Any singleton set $\{x\}$ is a terminal object, and the empty set \emptyset is the initial object. The truth-object is the set $\{\text{true}, \text{false}\}$ (denoted by 2), and $\text{true}: 1 \rightarrow 2$ is the map $x \mapsto \text{true}$ for any singleton $\{x\}$. An exponential object B^A is the set of functions from A to B , denoted by $\underline{\text{Set}}(A, B)$. Given a function $g: C \times A \rightarrow B$, there is an exponential adjoint $\hat{g}: C \rightarrow B^A$, given by $c \mapsto g_c$, such that $\text{ev}: B^A \times A \rightarrow B$ is given by $\text{ev}(\hat{g}(c), a) = g_c(a) = g(c, a)$. It is well-known that there is an equivalence between subsets (= monics, = inclusions) $m: A \rightarrow X$ and characters $\text{char}(m): X \rightarrow 2$, that is the powerset $P(X)$ is equivalent to 2^X . We have a string of equivalences

$$P(1) \cong 2 \cong 2^1 \cong \underline{\text{Set}}(1, 2).$$

$\underline{\text{Set}}$ is, perhaps, the paradigm example of a topos.

Nevertheless, there are many other categories satisfying the axioms of topos theory.

Proposition 3 Any category, $\underline{\text{Set}}^{\underline{C}^{op}}$, of variable sets (dually $\underline{\text{Set}}^{\underline{C}}$) is a topos.

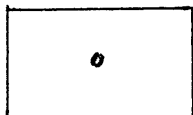
Proof : Goldblatt ([GB] pp.204-10).

Certainly E1 should hold since limits and colimits are inherited

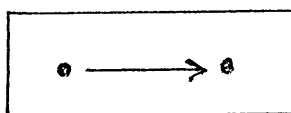
pointwise from Set. Adopting the fruitful confusion between an object A in \underline{C} and the representable functors h_A (Chap.5, Prop.5), then the Yoneda Lemma (Chap.5, Prop.4) tells us that, for any presheaf F , the natural transformations from A to F are in 1-1 correspondence with the set $F(A)$ defined at stage A . This can be used to determine exponentiation in variable sets. If F and G are variable sets, then if F^G is to exist at all then by E2 we must have

$$\frac{A \dashrightarrow F^G}{A \times G \dashrightarrow F} .$$

We use this to define F^G , by letting the maps $A \times G \dashrightarrow F$ be the elements of F^G at stage A . For example, working in RGraph, the Yoneda functor enables us to picture 1 as



and I as

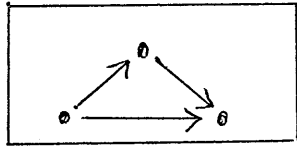


F^G at stage 1 (= nodes) is given by the set of graph morphisms from $1 \times G$ to F . F^G at stage I (= edges) is given by the graph maps from $I \times G$ to F . The cohesion of the exponential graph F^G is given by the restriction maps $F^G(h): F^G(I) \dashrightarrow F^G(1)$, where h is either $s: 1 \dashrightarrow I$ or $t: 1 \dashrightarrow I$. We have

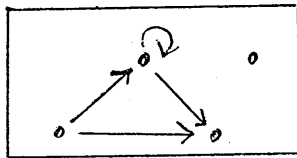
$$\begin{array}{ccc} & 1 \times G & \\ h \times \text{id}_G \swarrow & & \searrow n \\ I \times G & \dashrightarrow & F \\ & e & \end{array}$$

and if n is a node in F^G and e is an edge, then the restriction of e along $F^G(h)$ identifies that n , such that $n = e.(h \times \text{id})$.

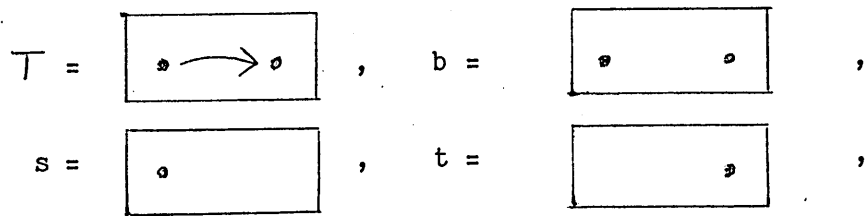
Thus we can picture the exponential object I^I in RGraph as



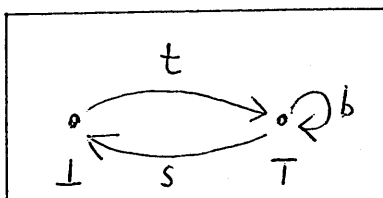
Note, however, that I^I must be pictured as



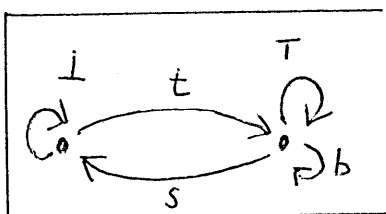
in Graph. There are no major difficulties in showing that exponential adjoints $\hat{g}: C \dashrightarrow B^A$ exist for any $g: C \times A \dashrightarrow B$, as calculations are done pointwise as in Set. Thus Set $_{\subseteq^{op}}$ is cartesian closed. To show that Set $_{\subseteq^{op}}$ satisfies axiom E3, we define Ω as equivalent to Ω^1 as in Set. By Yoneda the exponential object Ω^1 has natural transformations $A \dashrightarrow \Omega^1$ as elements of $\Omega(A) (\cong \Omega^1(A))$ at stage A for a representable A. However, these elements are in 1-1 correspondence with a morphism $1 \times A \dashrightarrow \Omega$, but these are just the maps from A to Ω , as $1 \times A \cong A$ in a cartesian closed category. However, according to E3 characters $h: A \dashrightarrow \Omega$ are in 1-1 correspondence with monics $\ker(h): R \dashrightarrow A$, so that R is a subobject of A. We can use this to define $\Omega(A)$. At stage A, $\Omega(A) \cong \{R \mid R \text{ is a subobject of } A\}$. For example, in RGraph at stage 1, the only subgraphs of 1 are 1 itself ($= \top$) and the empty graph ($= \perp$). At stage I the subgraphs of I are



and the empty graph ($= \perp$). Putting stages 1 and I together with the cohesion of restricting along the source and target, we can picture the truth-graph as



The truth-graph in Graph is the same graph as in RGraph, but must be pictured as



Clearly (in RGraph) we can define $\text{true}: 1 \longrightarrow \Omega$ as the map that sends $\boxed{\bullet}$ to T in Ω , and $\text{true}_A: A \longrightarrow \Omega$ for any graph A sends all nodes to T and all edges to the degenerate loop attached to node T . Furthermore, the truth-graph is the universal object required by axiom E3 to classify the subgraphs A of X , since the 'inclusion' of a subgraph A in a graph X brings about a unique character $X \longrightarrow \Omega$ (which is a graph homomorphism) in the following way :

- a) all nodes of X in A must go to node T ;
- b) all edges of X in A must go to the degenerate loop at T ;
- c) all nodes of X not in A must go to node I ;
- d) all edges of X not in A must go to the degenerate loop at I ,
 t , s , or b , depending on whether both their source and target are not in A , only their target is in A , only their source is in A , or both their source and target is in A .

No other graph can serve this role. For example, if the edge b were to be removed then only 'full' subgraphs could be classified. Thus all of Ω is needed because of the variety of 'inclusions' which exist in categories of graphs. Given any homomorphism $h: X \longrightarrow \Omega$,

it is easy to see that the domain of $\ker(h)$ is isomorphic to the subobject classified by h .

Proposition 4 (Fundamental Theorem of Topoi).

If \underline{E} is a topos then the slice category \underline{E}/X , where X is any object of \underline{E} , is a topos.

Proof : ([FR] Theorem 2.31, pp.24-6).

Without going into the complexities of the proof, we know that for presheaves, $\underline{\text{Set}}^{\underline{C}^{op}}/X$ is equivalent to $\underline{\text{Set}}^{\underline{X}^{op}}$, where \underline{X} is the discrete fibration associated with the object X (Chap.5. Prop.7,8). Since \underline{X} is a small category then by Proposition 3 then $\underline{\text{Set}}^{\underline{X}^{op}}$ is a topos.

The previous propositions (3 and 4) provide us with a plenitude of topoi that share considerable similarities to $\underline{\text{Set}}$. One could go on to add further axioms of infinity, Booleanness, choice, and that Ω be bivalent ([TY]). In this way, one could arrive at a theory of the category of sets (expressed in the language of category theory) such that (upto equivalence) $\underline{\text{Set}}$ was the only model ([LWS,MI,TY]). However, axioms E1 to E3 are just good enough to treat topoi as set-like categories for many purposes, including the treatment of elements, functional completeness, set-theoretic operations on objects, and (most importantly) as the target of an interpretation of a type theory.

Sets have elements, and they can always be represented by a function $x:1 \rightarrow X$ ('select the element x '). This leads us to the following.

Definition 4 An element (or point, or global element) in a topos is a section $x:1 \rightarrow X$. More generally, a generalised element is

an arrow with codomain X . The domain of the arrow is called the stage of definition of the element. A generic element in a variable set is an arrow $A \dashrightarrow X$, where A is a representable (= a generator).

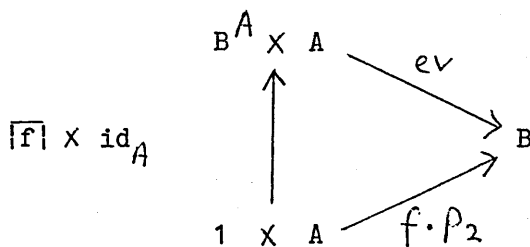
Thus the fact that the only generic elements in a constant set are the global sections derives from Set having 1 as its only generator. Variable sets are abstractly the same as Set in that the generic elements are derived from the representables. To be sure, we must distinguish points from generic elements.

Definition 5 An object X in a topos \mathcal{E} is called non-zero if it is not isomorphic to the initial object, and non-empty if there is at least one global element $1 \dashrightarrow X$. \mathcal{E} is said 'to have empty objects' if any non-zero objects are empty.

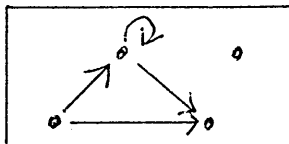
Thus Set does not have empty objects; its only elements are the global elements, and non-zero and non-empty are coextensive. RGraph does not have empty objects as 1 is a representable. However, the situation in Graph is somewhat different. It is only too easy to have irreflexive graphs without global elements, and this is clearly related to the fact that 1 is not in the set of generators for Graph. Thus Graph exemplifies those topos with empty objects. Graph has nodes and edges as generic elements, whereas RGraph has points and edges. The secret of variable sets would appear in choosing a topos with the generic elements that one needs.

The fact that topos are cartesian closed means that any topos has the following important analogues to similar constructions in Set.

1) Consider in Set, the following diagram



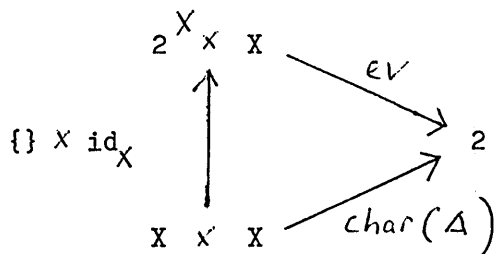
where given $f:A \rightarrow B$, $[\overline{f}]:1 \rightarrow B^A$ is the name of f , and is an exponential adjoint. $[\overline{f}]$ picks out f in $\{g|g:A \rightarrow B\}$ ($= B^A$), such that the above diagram commutes. Recall (Chap.6 Def.8) that $[\overline{f}]$ picks out $(\lambda x \in A) \phi(x)$ which is the rule $x \mapsto \phi(x)$ defining f . Now these diagrams exist in any topos \underline{E} , and this implies \underline{E} is just good enough to provide the codomain of an interpretation with function types and closure under application. For example, in RGraph $[\overline{f}]$ must pick out at stage C a graph map $C \times A \rightarrow B$, where C is a generic element. It is obvious at stage 1, $[\overline{f}]$ picks out f in $\{g|g:A \rightarrow B\}$ where g is a graph map. This works with empty objects too. For example, in Graph I is an empty object, and there is only one (irreflexive) graph homomorphism ($= id_I$) from I to I . If we picture the object I^I as



then the name of that unique homomorphism is the map

$[\overline{id_I}]:1 \rightarrow I^I$ that picks out the node with the only loop.

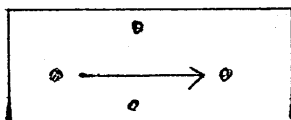
2) Consider in Set, the following diagram



where $char(\Delta)$ is the character of

$$\langle \text{id}_X, \text{id}_X \rangle : X \longrightarrow X \times X \quad (x \longmapsto \langle x, x \rangle)$$

and $\{\}$ is the exponential adjoint. The function $\{\}: X \longrightarrow P(X)$ is the singleton map sending an element x to $\{x\}$. Analogues of this situation exist in any topos. Thus if $P(X)$ is defined as Ω^X , then there exists a singleton map $\{\}: X \longrightarrow P(X)$. In RGraph the singleton map at stage 1 sends a node x in X to that graph map $1 \times X \longrightarrow \Omega$ in Ω^X such that node x takes the value \top and all other nodes take the value \perp . An edge e in X is sent to a graph map $I \times X \longrightarrow \Omega$ in Ω^X , such that e takes the value \top and other edges take appropriate values depending on how their source targets are related to the source and targets of e . Thus the concept of singletons finds an appropriate expression in any topos. For example, in Graph, the powerset $P(I)$ is the graph Ω^I . This has 4 nodes and 20 edges and is too elaborate to picture here. However, the singleton map $\{\}: I \longrightarrow P(I)$ can be approached in the following way. If $I \times I$ is pictured as



then the singleton at stage I sends I to the map $I \times I \longrightarrow \Omega$ which sends the only edge and its source and target to \top and the isolated nodes to \perp .

In some ways the truth-object in a topos is rather like an injective space in topology.

Definition 6 An injective in a category is an object I such that for every monic $m: B \longrightarrow A$ and every arrow $f: B \longrightarrow I$ there is an arrow $g: A \longrightarrow I$ such that $gm = f$. Dually, a projective is an object P such that for every epi $e: A \longrightarrow B$ and every arrow $f: P \longrightarrow B$ there is an arrow $g: P \longrightarrow A$ such that $eg = f$.

In general it is a difficult task to sort out the injectives and projectives in a category. But clearly 1 is projective and injective in RGraph. 1 is not projective in Graph (hint: let $B = 1$). The following clarifies the status of Ω .

Proposition 5 In a topos Ω is an injective and Ω^A is an injective for any object A .

Proof : ([FR] Prop.2.52, 2.51 pp.31-2).

Essentially, this means that the singleton map $\{ \} : X \rightarrow P(X)$ embeds X into an injective object.

As an example of how one operates with Ω in a topos, consider the algebraic structure Ω must carry.

a) We already have a map $\text{true} : 1 \rightarrow \Omega$, which is readily seen as the character of $\text{id}_1 : 1 \rightarrow 1$.

b) Similarly, after establishing $0_1 : 0 \rightarrow 1$ is monic, we have $\text{false} : 1 \rightarrow \Omega$ as its character.

c) We can establish conjunction as an arrow

$$\wedge : \Omega \times \Omega \rightarrow \Omega$$

which is the character of the product arrow

$$\langle \text{true}, \text{true} \rangle : 1 \rightarrow \Omega \times \Omega.$$

d) Similarly disjunction $\vee : \Omega \times \Omega \rightarrow \Omega$ which is the character of the image of the coproduct arrow

$$[\langle \text{id}_\Omega, \text{true}_\Omega \rangle, \langle \text{true}_\Omega, \text{id}_\Omega \rangle] : \Omega + \Omega \rightarrow \Omega \times \Omega.$$

e) Material implication $\Rightarrow : \Omega \times \Omega \rightarrow \Omega$ which is the character of $e : \text{Ker}(\Rightarrow) \rightarrow \Omega \times \Omega$ where e is the equaliser of $\Omega \times \Omega \xrightarrow[\rho_1]{\wedge} \Omega$. It will be recognised that the domain of e is the partial ordering of elements of Ω .

f) Negation is $\neg : \Omega \rightarrow \Omega$ which is the character of $\text{false} : 1 \rightarrow \Omega$.

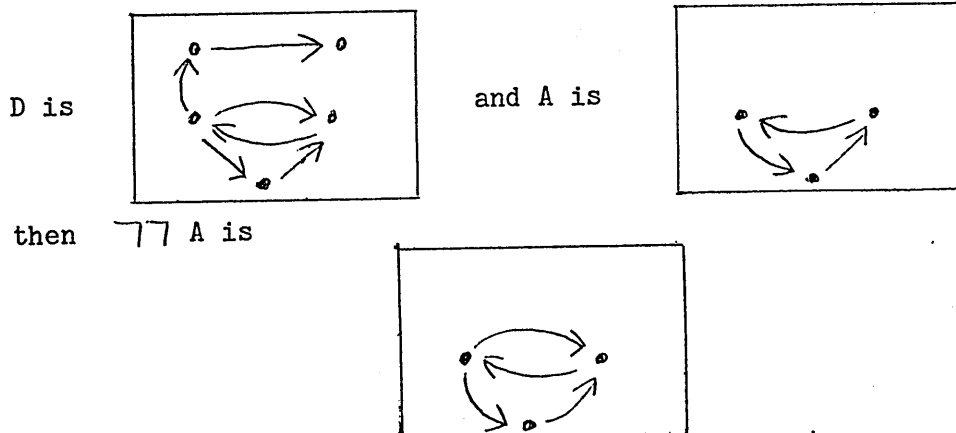
Proposition 6 Ω is a Heyting algebra object, with minimum \perp given by the image of false: $1 \rightarrow \Omega$, maximum \top given by the image of true: $1 \rightarrow \Omega$, and \wedge , \vee , and \Rightarrow as defined above.
 Proof : ([BW] p.197).

The reader will recall that a Heyting algebra is a Boolean algebra if $(\forall t \in \Omega) \neg \neg t \Rightarrow t$. The Boolean algebras form a subclass of Heyting algebras.

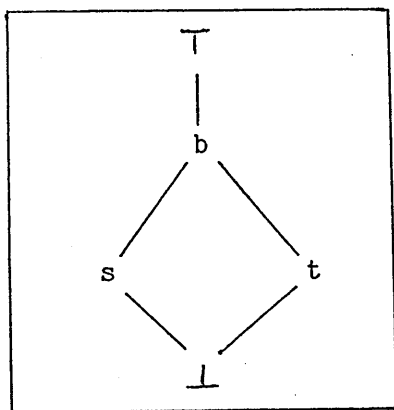
The analogues of set-theoretical constructions are now available in any topos. Let D be an object in a topos, then we can define operations on the collection of subobjects of D as follows.

- a) Complements : given a monic $f: A \rightarrow D$, the complement of f relative to D is the kernel of $\neg \circ \text{char}(f)$.
- b) Intersections : the intersection of monics $f: A \rightarrow D$ and $g: B \rightarrow D$ is the subobject $A \cap B \rightarrow D$ which is the kernel of $\wedge \circ \langle \text{char}(f), \text{char}(g) \rangle$.
- c) Unions : the union of the monics mentioned above is the monic $A \cup B \rightarrow D$ with $\vee \circ \langle \text{char}(f), \text{char}(g) \rangle$ as its character.

However, these set-theoretical operations may differ somewhat from ordinary set operations. It is well-known that the algebra of subsets is Boolean. In particular, $\neg \neg A$ in D is equivalent to A . Underscoring the intuitionist character of topos theory, we have the fact that $\neg \neg : \Omega \rightarrow \Omega$ is rarely the identity in a topos. For example, in RGraph if



In general, $\neg\neg A$ is that subobject whose elements are, on restriction, in A and $A \subseteq \neg\neg A$ but $\neg\neg A$ is not contained in A . Once it is recognised that $\Rightarrow: \Omega \times \Omega \dashrightarrow \Omega$ induces containment then, in general, we can write $A \Rightarrow \neg\neg A$ but not $\neg\neg A \Rightarrow A$. Thus the algebra of subgraphs is not classical. At stage I the truth-graph only carries the structure of a Heyting algebra



in which it can be seen that $\neg\neg b = T$ is its only departure from Booleanness. Indeed RGraph is an example of the situation in which "... no logic stronger than intuitionistic logic can be valid for sets that are varying in any serious way" ([LWV] p.105). The nature of the truth-object can be used to characterise an important class of topoi.

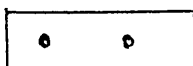
Definition 7 A topos \underline{E} will be called Boolean iff Ω is a Boolean algebra, that is $(\forall t \in \Omega) \neg\neg t \Leftrightarrow t$.

Proposition 7 For any topos \underline{E} , the following are equivalent :

- a) \underline{E} is Boolean;
- b) $[true, false]: 1 + 1 \dashrightarrow \Omega$ is iso, that is $1 + 1 \cong \Omega$.

Proof : ([GB] pp.156-7).

It is readily seen that $1 + 1$ in RGraph, pictured as



is not isomorphic to the truth-graph. This proposition provides us

with a clear test of Booleaness. Once it is recognised that the Boolean algebras are a subclass of the Heyting algebras, we see that the manipulation of constant sets is just a special case of the manipulation of subobjects in a topos. A topos is just good enough for naive intuitionistic set theory, and a topos theorist's abuse of describing objects as sets has considerable justification.

The most important aspect of topoi, however, is that they are just good enough to serve as the codomain of an interpretation. On the one hand, every topos \underline{E} has an internal language, denoted by $L(\underline{E})$. This is constructed in exactly the same way as $L(\underline{Set})$ (Chap.6, Def.8). To be sure, the word 'object' should be substituted for the word 'set', and the types 1 and Ω are the objects 1 and Ω . Functions must be called morphisms. Where I have described characters as taking the value 'true, false otherwise', we must take account of the variety of 'inclusions' Ω permits by saying ' \top ', or some other appropriate value of Ω '. Nothing precludes a topos with empty objects having an internal language. Empty objects become empty types in $L(\underline{E})$. On the other hand, every type theory A generates a topos $\underline{T(A)}$. The details of this will not be needed. However, the importance of the topos generated by A can be assessed by the following.

Definition 8 The category Lang has type theories A as objects. The morphisms are translations. Translations preserve $1, \Omega$, powerset and product formation, and the closure of function types. They send closed terms to closed terms of corresponding types so as to preserve $*$, \in , $\{ \}$, and $\langle \rangle$ upto provable equality. They send variables to variables in a prescribed way : with the i 'th variable in the domain being sent to the i 'th variable in the codomain. They

send theorems to theorems. For more details see ([LS] p.197).

Definition 9 The category Topoi has topoi as objects. Morphisms are logical functors which preserve $1, \bigcap, \bigcup$, powersets, products, and exponentials.

Proposition 8 There is a functor $L: \text{Topoi} \rightarrow \text{Lang}$ which sends every topos \underline{E} to its internal language $L(\underline{E})$, and a functor $T: \text{Lang} \rightarrow \text{Topoi}$ which sends every type theory A to the topos, $T(A)$, generated by it. Furthermore, T is left adjoint to L , and the counit $\varepsilon(\underline{E}): TL(\underline{E}) \rightarrow \underline{E}$ is an equivalence of categories, that is $TL(\underline{E}) \cong \underline{E}$.

Proof : ([LS] Theorem 15.4, 15.5 pp.204-5).

Actually the above proof only works for a subcategory of Topoi of 'topoi with canonical subobjects with strict logical functors'. But as every topos \underline{E} is equivalent to one with canonical subobjects, namely $TL(\underline{E})$ ([LS] pp.201-2), and since the variable sets which interest Science are such topoi 'with canonical subobjects', then we ignore these technical complexities. To give details of this adjunction would take us too far afield. What is important is the uses to which it can be put. Given that $T \dashv L$, we have the scheme

$$\frac{\text{Topoi}(T(A), \underline{E})}{\text{Lang}(A, L(\underline{E}))},$$

in which the counit evaluates as an equivalence. Thus Lambek and Scott point out that ([LS] pp.123-4) :

"While not every type theory is the internal language of a topos, every topos is equivalent to one generated by a type theory ... For us, an interpretation of a type theory A in a topos \underline{E} is a morphism $A \rightarrow L(\underline{E})$ in Lang or, in view of adjointness, a morphism $T(A) \rightarrow \underline{E}$ in Topoi".

So although only some type theories are internal languages of topoi, every type theory can be interpreted into the internal language of some topos. Previously I referred to an interpretation as a mapping $[]: H \rightarrow \text{Set}$. Strictly speaking, an interpretation

is a translation in Lang from H into $L(\underline{\text{Set}})$. But since the topos generated by $L(\underline{\text{Set}})$ is equivalent to Set then we are justified in taking the cruder approach of a mapping into Set. What Proposition 8 tells us is that, in principle, we can interpret our theories into other topoi, and given that topoi need not be Boolean then they are potential targets for an interpretation of an intuitionistic type theory.

The interpretation of a type theory into a topos of variable sets is, in principle, not different from that of an interpretation into Set. Notice that the entailment relation for Set (Chap.6, Def.9) was defined in such a way as to take no advantage that Set does not have empty objects. In classical logic, the presence of empty types would correspond to the presence of closed terms which do not always have a denotation (so-called free logic). However, it has long been recognised by topos theorists that one can obtain a set-like predicate logic if one quantifies not over the global elements of an object but rather over its generalised elements. Recall that entailment means that for all arrows $h:C \rightarrow A$ and for all sets C where $f_i:A \rightarrow \Omega = (\lambda x \in A) \phi_i(x)$ and $f_i.h = \text{true}_C$ then $f_{n+1}.h = \text{true}_C$. Since entailment is defined only in terms of generalised elements, this syntactical approach can be used for variable sets too, with the topos-theoretic abuse that objects are just 'sets'. However, it may be more illuminating to recast this syntactical approach in terms of interpretation vis-à-vis stages of definition through what is called 'Kripke-Joyal Semantics'.

Definition 10 If $\phi(x)$ is a formula in $L(\underline{E})$ in the variable x of type A then $\phi(x) = f(x)$ for a uniquely determined arrow $f:A \rightarrow \Omega$. For an arrow $a:C \rightarrow A$ in \underline{E} we write $\phi(a)$ is $f.a$,

by abuse of notation, and regard $a:C \rightarrow A$ as a generalised element of A at stage C . We write $C \Vdash \phi(a)$ for $f.a = \text{true}_C$, and say ' $\phi(a)$ holds at stage C ' or ' C forces $\phi(a)$ '. More generally, in the presence of a parameter y of type B , if $\phi(y,x) =_{\{x,y\}} g\langle y,x \rangle$ where $g:B \times A \rightarrow \Omega$ then we write $\phi(y,a)$ is $g.\langle y, a \rangle$.

The following are consequences of Definition 10.

Proposition 9 If $C \Vdash \phi(a)$ and $h:D \rightarrow C$ then

$D \Vdash \phi(a.h)$.

Proof sketch: If $C \Vdash \phi(a)$ then by Definition 10 we have $f.a = \text{true}_C$. Given $h:D \rightarrow C$ and composing with a , we must have $f.a.h = \text{true}_D$. Since $\text{true}_C.h = \text{true}_D = f.a.h$ then this means that $D \Vdash \phi(a.h)$.

Proposition 10 $\vdash x \phi(x)$ in $L(\underline{E})$ iff for all objects C and all generalised elements $a:C \rightarrow A$ then $C \Vdash \phi(a)$.

Proof sketch: $\vdash x \phi(x)$ means (Chap.6, Def.9-10) that for all objects C and all arrows $a:C \rightarrow A$ that $f.a = \text{true}_C$ for a uniquely determined arrow $f:A \rightarrow \Omega$; that is, by Def.10,

$C \Vdash \phi(a)$. Conversely, suppose $C \Vdash \phi(a)$ for all objects C and all $a:C \rightarrow A$, then $f.a = \text{true}_C = \text{true}_A.a$ for all C and all $a:C \rightarrow A$, and so $f = \text{true}_A$; that is $\vdash x \phi(x)$.

Proposition 11 $C \Vdash \phi(a)$ iff $\vdash (\forall z \in C) \phi(a.z)$.

Proof sketch: $C \Vdash \phi(a)$ means $f.a = \text{true}_C$; that is, $f.a.z = \top$, which translates to $\vdash z \phi(a.z)$, or (Chap.6, Def.5(8)) $\vdash (\forall z \in C) \phi(a.z)$. Conversely, suppose $\vdash z \phi(a.z)$, then by Proposition 10 $C \Vdash \phi(a)$.

Proposition 12 If $h:D \rightarrow C$ is epi and $D \Vdash \phi(a.h)$ then $C \Vdash \phi(a)$.

Proof sketch: $D \Vdash \phi(a.h)$ means that $f.a.h = \text{true}_D$. Now suppose $C \Vdash \phi(a)$ then $f.a = \text{true}_C$ and $\text{true}_C.h = \text{true}_D$. The definition of an epi is that it is right-cancellable (Chap.5, Def.10). So if h is epi then $f.a.h = \text{true}_C.h$ implies $f.a = \text{true}_C$. The latter simply means $C \Vdash \phi(a)$.

Proposition 13 If G is a generating set of objects of \underline{E} , then $\Vdash x \phi(x)$ in $L(\underline{E})$ iff for all objects C in G and all generalised elements a of A at stage C , then $C \Vdash \phi(a)$.

Proof sketch: Suppose $C \Vdash \phi(a)$ for all C in G and all $a: C \rightarrow A$ then $f.a = \text{true}_C = \text{true}_A.a$ for all C and a , and therefore $f = \text{true}_A$; that is $\Vdash x \phi(x)$. By Proposition 10, the converse is evident.

The meaning of Propositions 9-13 is clear. By Proposition 10, truth in a topos is equivalent to truth at all stages and for all generalised elements and this conforms with the earlier treatment of entailment. Proposition 13 allows us to restrict the stages of definition to a generating set. Proposition 1 means we can restrict these to the representables. So working in RGraph, Proposition 9 means that truth forced at stage I (= edges) implies truth is forced along the restriction of source and target maps to truth at the later stage of 1 (= points). Proposition 12 means that truth forced at loops implies truth is forced at points. Proposition 11 means that truth forced at any stage implies truth is forced at all later stages. This leads to 'Kripke-Joyal Semantics'.

Proposition 14 (For variable sets)

Given a generic element $h: C \rightarrow A$ defined at stage C , then :

- 1) $C \Vdash \phi(a)$ iff $a = \text{true}_C$, when $A = \Omega$.
- 2) $C \Vdash b \in S$ iff $\text{ev}(S, b) = \text{true}_C$ when $a: C \rightarrow P(A) \times A$ is $\langle S, b \rangle: C \rightarrow P(A) \times A$.

- 3) $C \Vdash \top$ always.
- 4) $C \Vdash \perp$ never.
- 5) $C \Vdash \phi(a) \wedge \psi(a)$ iff $C \Vdash \phi(a)$ and $C \Vdash \psi(a)$.
- 6) $C \Vdash \phi(a) \vee \psi(a)$ iff $C \Vdash \phi(a)$ or $C \Vdash \psi(a)$.
- 7) $C \Vdash \phi(a) \Rightarrow \psi(a)$ iff for all $h:D \rightarrow C$ if $D \Vdash \phi(a.h)$ then $D \Vdash \psi(a.h)$.
- 8) $C \Vdash (\forall y \in B) \psi(y,a)$ iff for all $h:D \rightarrow C$ and all $b:D \rightarrow B$ then $D \Vdash \psi(b,a.h)$.
- 9) $C \Vdash (\exists y \in B) \psi(y,a)$ iff there is an epi. $h:D \rightarrow C$ and an arrow $b:D \rightarrow B$ such that $D \Vdash \psi(b,a.h)$.
- 10) $C \Vdash \neg \phi(a)$ iff for all $h:D \rightarrow C$ if $D \Vdash \phi(a.h)$ then $D \approx 0$.

Proof : ([LS] pp.166-8, 172-3).

The above constitutes the useful logical machinery for conceiving models of any theory with an interpretation in variable sets.

For example, a model of ThC in RGraph would mean that the axioms of category theory would have to be forced at all stages of definition. Thus models of ThC in RGraph would be graphs in which both the sets of points and edges were small categories and in which the restriction maps would be functors between them. Obviously a model of ThC in RGraph consists of two graphs corresponding to the types of arrows and objects. So we have as a category object in RGraph, a graph of arrows and a graph of objects with graph homomorphisms between them for domain, codomain, and identity mappings. (Indeed many of our results on variable sets could be extended to such topoi as $\underline{E}^{\underline{C}^{\text{op}}}$ where \underline{C} is a category object in \underline{E}). Again this illustrates a general approach. If H is some theory about mathematical objects, then an interpretation in a

topos \underline{E} produces an H-object in that topos. Thus we can have natural number, real number, lattice, group, ring, metric, and topological space objects in \underline{E} . Briefly we may say that "... the notion of a topos summarizes in objective categorical form the essence of 'higher order logic' ... with no axiom of extensionality. This amounts to a natural and useful generalization of set theory to the consideration of 'sets which internally develop'" ([LWI] p.3). The working mathematician will realise that the category of sets admits finite limits and powerset constructions. These are the essential tools to carry out all mathematical constructions and these are the features of sets that are generalised to topoi .

In what sense are interpretations valid ? If A is a type theory and $[]:A \longrightarrow \underline{E}$ is an interpretation into \underline{E} then, by abuse, we write $[]:A \longrightarrow L(\underline{E})$ as the translation corresponding to an interpretation, and we call this a model (or interpretation) too. A standard model is a translation $[]:A \longrightarrow L(\text{Set})$, and for a provable formula p in A we have $L(\text{Set}) \vdash [p]$ iff $\text{Set} \vdash p$ as the Completeness Theorem (Chap.6, Prop.4). Our discussion of Taylor's formula in differential calculus with Berkeley's attack on Newton and Leibniz shows that, in general, there are not enough standard models. Among all the interpretations of A there is the canonical one, given by the unit $\gamma(A):A \longrightarrow LT(A)$, and we have $LT(A) \vdash [p]$ iff $T(A) \vdash p$. Furthermore, for a formula provable by Kripke-Joyal Semantics we have $L(\underline{E}) \vdash [p]$ iff $\underline{E} \vdash p$ ([OS] p.351). If we admit all interpretations then the Completeness theorem holds trivially, and the unit of $T \dashv L$ is initial in the category of all interpretations. The super-logical reader might look for a notion of a model $[]:A \longrightarrow L(\underline{M})$, in which the topos \underline{M}

resembles Set more closely than an arbitrary topos.

Definition 11 A model of type theory A is a translation

$[]:A \longrightarrow L(\underline{M})$, in which the topos \underline{M} (also confusingly called a 'model') has the following properties :

M1) $\text{Not } \underline{M} \models \perp$ (no contradiction holds in \underline{M}).

M2) if $\underline{M} \models p \vee q$ then either $\underline{M} \models p$ or $\underline{M} \models q$

(disjunctive property).

M3) if $\underline{M} \models (\exists x \in A) \phi(x)$ then there is an entity a of type A (actually an arrow $a:1 \longrightarrow A$) such that $\underline{M} \models \phi(a)$

(existence property).

The disjunctive property means that 1 is indecomposable; that is, 1 is not the union of two proper subobjects. The existence property means that 1 must be projective ([LS] p.213). Property M1 is only there to rule out the trivial topos, 1 , in which 1 is also initial. For every model \underline{M} there is a left exact functor $\gamma:\underline{M} \longrightarrow \text{Set}$, which sends an object to its set of global elements. When \underline{M} is Boolean then γ is faithful and we may regard \underline{M} as a subcategory of Set ([LS] p.213). Indeed, the import of Henkin's work on completeness is that when \underline{M} is Boolean then there are enough non-standard models ([LS] Theorem 17.6, p.216). Such non-standard models form 'the actual advances that set theorists and logicians have made' (e.g. Cohen on the Continuum Hypothesis and Robinson on Non-standard Analysis [DH]). When \underline{M} is not Boolean then \underline{M} is a model of intuitionistic set theory. It is easy to see that RGraph is such a model, but Graph is not as 1 is not projective. The general completeness theorem for higher order logic asserts :

Proposition 15 Given any type theory A , $L(\underline{M}) \vdash [p]$ for a provable formula p in A iff $\underline{M} \models p$ for a topos \underline{M} satisfying Definition 11 and all translations $[]:A \longrightarrow L(\underline{M})$.

Proof : ([LS] Theorem 19.1 pp.223-5).

In the course of this proof, it is established that the disjunctive and existence properties are metatheorems of intuitionist type theory. Working mathematicians are often content if a formula is satisfiable, that is, valid in at least one model. The emphasis is on the 'showing'. However, if one wants more from set theory and logic then one can always restrict attention to topoi satisfying Definition 11.

Given that type theories can be interpreted in a topos, then a fairly traditional way to proceed is to formulate some conception of 'space' in type theory and to interpret this in some topos to exhibit the 'space' objects therein. However, this ignores the context in which the notion of a topos was originally introduced by Grothendieck as a natural generalisation of that of a topological space ([JT3]). Since no discussion of topoi would be complete without reference to these topological notions, I give a brief outline of some of these ideas to offset them from those cases in which a topos actually is a category whose objects are spaces.

A useful starting point is the notion of a geometric morphism. Recall that for every continuous map $f:X \rightarrow Y$ between topological spaces, it holds that the inverse image of $f^{-1}:O(Y) \rightarrow O(X)$ preserves finite limits and arbitrary suprema of the lattice of open sets. Related to $f:X \rightarrow Y$ there is also another meet and suprema preserving function $f_*:O(X) \rightarrow O(Y)$ $[U \rightarrow \text{int}(Y - f(X-U))]$ which can be seen as a right adjoint to f^{-1} , when the complete Heyting algebra of open sets of a space is treated as a small category. To generalise this to topoi, we have:

Definition 12 Let $\underline{E}, \underline{F}$ be topoi. A geometric morphism $f: \underline{F} \rightarrow \underline{E}$ consists of a pair (f^*, f_*) of functors $f_*: \underline{F} \rightarrow \underline{E}$, $f^*: \underline{E} \rightarrow \underline{F}$ (called the direct and inverse images of f) such that $f^* \dashv f_*$ and f^* is left exact. \underline{F} is sometimes called an \underline{E} -topos. A geometric morphism is said to be essential if f^* also has a left adjoint $f_!: \underline{F} \rightarrow \underline{E}$. f is called an inclusion if f_* is full and faithful, and a surjection when f^* is faithful.

Proposition 16 If $f: \underline{E} \rightarrow \underline{\text{Set}}$ is a geometric morphism then it is unique.

Proof : ([MI], [JT1] Prop.4.41 pp.119-20).

In fact f is the pair (Δ, γ) where γ is the 'points' functor which sends an object X to its set of points $\underline{E}(1, X)$. The inverse image $\Delta: \underline{\text{Set}} \rightarrow \underline{E}$ is the 'discrete' functor, sending a set X to the constant set $X(A)$ at each stage A ; that is, a set is sent to a discrete object in \underline{E} . It can readily be seen that the pair (Δ, γ) is a surjection.

The composition $f_* \cdot f^{-1}: \underline{O}(Y) \rightarrow \underline{O}(Y)$ has a number of interesting properties. Writing j for $f_* \cdot f^{-1}$, the endofunctor $j: \underline{O}(Y) \rightarrow \underline{O}(Y)$ has for U, V open in Y ([VR] p.23):

- i) $j(Y) = Y$;
- ii) $U \subseteq j(U)$;
- iii) $j.j(U) = j(U)$;
- iv) $j(U \cap V) = j(U) \cap j(V)$.

In topos theory, it turns out that an endomorphism on a complete Heyting algebra object satisfying these four properties is very fruitful. For example, the reflective subcategories (= subtopoi) of a topos are in 1-1 correspondence with such endomorphisms on its truth object.

Definition 13 A (Lawvere-Tierney) topology on Ω is a morphism

$j: \Omega \rightarrow \Omega$ such that (in the internal language, [VR] p.24):

i) $j \top = \top$;

ii) $(\forall t \in \Omega) j.jt = t$;

iii) $(\forall t, t' \in \Omega) j(t \wedge t') = j(t) \wedge j(t')$;

or more categorically :

a) $j.\text{true} = \text{true}$;

b) $j.j = j$;

c) $j.\wedge = \wedge .(j \times j)$.

There is a partial order on topologies. $j \leq j'$ iff $(\forall t \in \Omega) jt \leq jt'$, and j' is said to be finer than j .

For example, the following endomorphisms on Ω are topologies.

a) $\text{id}, t \mapsto t$ (minimal topology).

b) $\text{max}, t \mapsto \top$ (maximal topology).

c) $\neg\neg, t \mapsto \neg\neg t$ (double negation topology).

d) $j_p, t \mapsto (t \vee p)$ (closed topology).

e) $j^p, t \mapsto (p \Rightarrow t)$ (open topology).

The double negation topology is used in classical mathematics for a variety of problems, such as completion and compactification

([VR]). In the context of Kripke's modelling of intuitionist logic

with a topos $\text{Set}^{\underline{P}}$ (where \underline{P} is a partial order of possible

worlds), one way of reading $\neg\neg \phi$ is to say that we keep open

the possibility to prove ϕ , or 'it is cofinally the case that

ϕ '. Lawvere has suggested a geometric interpretation of $j\phi$ as

'it is j -locally the case that ϕ ' ([LWI]).

The notion of a Grothendieck topology is related to that of a Lawvere-Tierney topology :

Definition 14 A Grothendieck topology on Ω is a subobject $J \subseteq \Omega$, such that ([VR] pp.28-9)

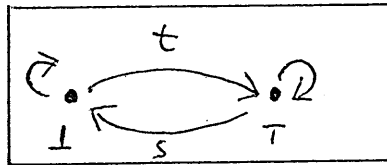
- i) $\tau \in J$;
- ii) $(\forall t \in J)(\forall t' \in \Omega)[(t \Rightarrow (t' \in J)) \Rightarrow (t' \in J)]$.

Proposition 17 The Lawvere-Tierney and Grothendieck topologies are in 1-1 correspondence.

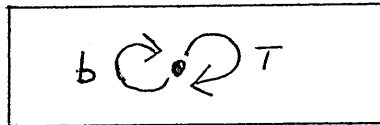
Proof : ([JT1] pp.32-3).

Thus to every Lawvere-Tierney topology j , there is a monic $h: J \rightarrow \Omega$ such that h is the kernel of j , and to every monic $h: J \rightarrow \Omega$ satisfying Definition 14 there is a Lawvere-Tierney topology $\text{char}(h)$ satisfying Definition 13.

For example, in Graph the double negation topology produces the graph $\Omega_{\neg\neg}$



as the image of $\neg\neg: \Omega \rightarrow \Omega$, and



is its corresponding Grothendieck topology.

Definition 15 Let $m: B \rightarrow A$ be a monic, with B^j as the kernel of $j.\text{char}(m)$, then ([VR] p.32)

- a) the j -closure B^j of B in A is the subtype

$$\{a \in A \mid j(\exists b \in B) m(b) = a\};$$

- b) B is j -closed if $B^j = B$;
- c) m is j -dense if $(\forall a \in A)j(\exists b \in B)[m(b)=a]$, that is B is dense iff $B^j = A$;
- d) A is a j -sheaf if

$$(\forall x \in A)[j(\exists ! a \in A)(a \in X \Rightarrow (\exists ! a \in A)j(a \in X))], \text{ that is,}$$

A is a j -sheaf iff for any dense monic $m: B \rightarrow A$ and for any $f: B \rightarrow C$ there is a unique $g: A \rightarrow C$, such that $g.m = f$.

A topology will be called standard iff all the representables are j -sheaves. The finest standard topology is called the canonical topology.

Proposition 18 Let j be a topology in \underline{E} , and $m:A \rightarrow X$ a monic, then m is j -dense iff $\text{char}(m)$ factors through $J \rightarrow \Omega$, and j -closed iff it factors through $\Omega_j \rightarrow \Omega$.

Proof : ([JT1] p.78).

This leads to the generalisation of the idea of a sheaf in classical mathematics. The following bears no resemblance to the usual definition in terms of covers, but is adequate for my purposes.

Definition 16 Let \underline{E} be a topos and denote the full subcategory of j -sheaves by \underline{E}_j . When \underline{E} is a category of presheaves $\text{Set}^{\underline{C}^{op}}$ then \underline{E}_j is called a Grothendieck topos, and the pair (\underline{C}, J) is called its site of definition, where J is the Grothendieck topology associated to j .

To be sure, all categories of presheaves are Grothendieck topos under the minimal topology, in which all presheaves are id-sheaves.

Unsurprisingly we have :

Proposition 18 A Grothendieck topos is an elementary topos, with the image Ω_j of $j:\Omega \rightarrow \Omega$ as its truth object.

Proof : ([JT1] pp.81-3).

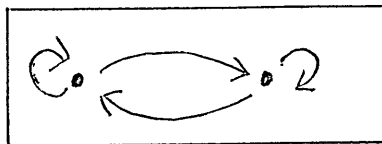
Whereas every Grothendieck topos is an elementary topos, not every elementary topos is a Grothendieck topos. For example, Finset, the topos of finite sets, is not a Grothendieck topos ([JT1] p.25). The motivation for using a Grothendieck topos is fairly simple : one tries to choose a topology in which the interesting objects are sheaves. To complete the analogy with sheaf theory, it can be shown that the inclusion $i:\underline{E}_j \rightarrow \underline{E}$ of j -sheaves into a category of

presheaves is the inverse image part of a geometric morphism with a right adjoint $L_j^* : \underline{E} \rightarrow \underline{E}_j$ called the associated sheaf functor, sending a presheaf to a sheaf. If A is a presheaf, then we construct its associated j -sheaf $L_j^*(A)$ as follows ([LWI] p.10).

1) Denote the image of the composite $A \xrightarrow{\{\}} \Omega^A \xrightarrow{j^*} \Omega_j^A$ by the symbol MA .

2) Let $\text{char}(m)$ be the character $\Omega_j^A \rightarrow \Omega$ of the inclusion $MA \subseteq \Omega_j^A$, then $L_j^*(A)$ is the kernel of the composite $j.\text{char}(m)$.

For example, using the double negation topology in Graph, the associated sheaf of $1 + 1$ can be pictured as



Since $L_{\neg\neg}(1+1) \cong \Omega_{\neg\neg}$ then it is clear that Graph $_{\neg\neg}$ is a Boolean topos (Prop.7). In fact this can be generalised.

Proposition 19 If \underline{E} is a Grothendieck topos then the subtopos $\underline{E}_{\neg\neg}$ of $\neg\neg$ -sheaves is a Boolean topos.

Proof : ([FR] Prop.2.62, p.38).

Thus the sense in which topoi might be called generalised spaces is captured by defining Top as the category (not a topos!) whose objects are Grothendieck topoi and whose morphisms are geometric morphisms. Notice that Top is a different category from Topoi. Whereas a logical functor in Topoi preserves all the logic, a geometric morphism is not a 'logical' arrow but a 'continuous' functor that preserves the fibrational structure of objects in a topos ([BN] pp.27-8). In fact, it can be shown that f^* preserves a fragment of logic known as coherent or geometric logic ([JT3]). Thus Top can be regarded as a generalisation of Sp, with

Grothendieck topoi replacing spaces and geometric morphisms replacing continuous maps.

In Top there is a certain confusion between objects and geometric morphisms. This is in line with the doctrine that adjunctions are the central concept of category theory. This confusion will be evident in the following notes on various 'topological properties' of geometric morphisms; the terminology being rather unsystematically borrowed sometimes from properties of spaces and sometimes from the properties of continuous maps. The following is an important sample of topological notions which have their analogues in Top.

Definition 16 A geometric morphism $f: \underline{F} \dashrightarrow \underline{E}$ is

- a) connected if f^* is full and faithful;
- b) hyperconnected iff the unit and counit of the adjunction $(f^* \dashv f_*)$ are both monic (or equivalently f_* preserves Ω);
- c) locally connected if f is essential (that is, f^* has a left adjoint $f_! : \underline{F} \dashrightarrow \underline{E}$);
- d) atomic (or smooth) if f^* is a logical functor;
- e) localic if every object of \underline{F} is a subquotient of one in the image of f^* (or equivalently there exists a complete Heyting algebra object A (= locale) in \underline{E} such that \underline{F} is equivalent to the canonical topology on $\underline{E} \text{--}\underline{A}^{cp}$).
- f) a local homeomorphism if there exists an object X in \underline{E} such that $\underline{F} \cong \underline{E}/X$.

The notion of connectivity is derived from the fact that if $\underline{Sh}(X)$ is the category of sheaves on a topological space X then the unique morphism $(\Delta, \gamma) : \underline{Sh}(X) \dashrightarrow \underline{Set}$ is connected iff X is a connected space ([JT3] p.84). This applies to the notion of local

connectivity too. Hyperconnectivity is a stronger form of connectivity. A localic topos is analogous to $\underline{\text{Sh}}(X)$ in the sense that a topological space X is a locale and $\underline{\text{Sh}}(X)$ is the subtopos of $\underline{\text{Set}}^{X^{op}}$ for the canonical topology. The 'smoothness' of 'atomic' comes from f^* preserving all the logic. It follows that if \underline{E} is Boolean then \underline{F} is a topos in which the lattice of subobjects of any object is a complete atomic Boolean algebra ([BD]). The notion of a local homeomorphism follows easily from the well-known connection between sheaves and local homeomorphisms, in which a continuous map $f:X \rightarrow Y$ induces a local homeomorphism $\underline{\text{Sh}}(X) \rightarrow \underline{\text{Sh}}(Y)$ iff it is a local homeomorphism in the usual sense.

Proposition 20 The following implications hold between the notions introduced in Definition 16 :

- a) connected \wedge atomic \Rightarrow hyperconnected \Rightarrow connected \Rightarrow surjective;
- b) local homeomorphism \Leftrightarrow atomic \wedge localic;
- c) atomic \Rightarrow locally connected.

Proof : ([JT3] p.86).

It is very easy to see that $(\Delta, \gamma) : \underline{\text{RGraph}} \rightarrow \underline{\text{Set}}$ is hyperconnected. It follows that $\underline{\text{RGraph}}$ is connected as a $\underline{\text{Set}}$ -topos, and (Δ, γ) is surjective. Notice that $\underline{\text{Graph}}$ is connected but not hyperconnected.

Proposition 21 Let \underline{F} be a $\underline{\text{Set}}$ -topos, then $\Delta : \underline{\text{Set}} \rightarrow \underline{F}$ has a left adjoint (that is, (Δ, γ) is essential) iff Δ preserves exponentiation.

Proof : ([BP] Theorem 2,15,16).

In the case of presheaves, it is easily seen that Δ preserves exponentiation since Δ is full and faithful, that is $\Delta(X^\gamma) \cong \Delta(X)^{\Delta(\gamma)}$. (In the case of topoi of j -sheaves, a necessary condition for the preservation of exponentiation would be that the

constant presheaves were also j -sheaves). It follows that both RGraph and Graph are locally connected. The left adjoint to Δ is the 'components' functor $\prod_0 : \underline{F} \rightarrow \underline{Set}$ which sends an object X to its set of connected components. In the case of graphs, this is just the connected components of a graph in the usual sense. More categorically $\prod_0 : \underline{Set}^{\underline{C}^{op}} \rightarrow \underline{Set}$ is given by coequalising the structural maps of an object X considered as a discrete fibration ([PA]). We have

$$X_1 \begin{array}{c} \xrightarrow{dom} \\ \xrightarrow{cod} \end{array} X_0 \xrightarrow{\quad} \prod_0(X)$$

as a coequaliser.

Nevertheless, "... the dictum 'a topos is a generalized space' is not entirely free from oversimplification" ([JT3] p.77). Whereas interpreting a theory of spaces formulated in type theory gives no guidance as to what constitutes a topos of spaces, it is also clear that the idea of a generalised topological space as it emerges in Top is too broad. How general should the idea of space be? In the case of 'sets', category theory proceeded with a strategy of axiomatising the category of sets to reveal a broad class of set-like categories. Can this strategy also be used to reveal a class of space-like categories appropriate to the needs of Science?

CHAPTER EIGHT

TOWARDS TOPOI OF SPACES

Reason has always existed, but not always in a rational form. Hence the critic can take his cue from every existing form of theoretical and practical consciousness and from this ideal and final goal implicit in the actual forms of existing reality he can deduce a true reality.

Marx to Ruge.

My argument, so far, has pursued two tacks. On the one hand, I have argued that, from a mathematical perspective, the most powerful representation of the Formal available to Science is type theory. In particular, I have remarked that even (what might be called) the simple experiments of Galileo require expressions in a type theory which is non-classical. Furthermore, through the adjunction established by Lambek (Chap.7, Prop.8), an interpretation of a valid type theory into some topos is possible. This is the passage between the Formal and the Conceptual that Science needs (Fig.2) to produce models for supply to Research Groups. On the other hand, I have argued that, from the perspective of the Philosophy of Science, the constitutive questions of a technical cognitive interest stem from a schematization of space, time, substance, and causality, in which natural phenomena are construed as non-intentional objects whose processes are to be explained in terms of a causal connection. The reduction of the logic of explanation to that of prediction requires the introduction of mathematics as a methodical device with natural

invariances described geometrically as elements of spaces.

Furthermore, secondary methodological objectifications constituting the practical interest permit the introduction of mathematics into the Human Sciences, either in terms of conditional predictions of quasi-causal processes or as a reference to 'structure' in some good-reason-assay. The time has now arrived to steer a direct course, in which the arguments of Philosophy and Mathematics are now more closely linked.

There is a certain indeterminacy in the injunction to interpret a type theory (even when it has no standard models) into some topos. On the one hand, there may well be theories which have an interpretation into every non-trivial topos, particularly those topoi which are models of set theory (Chap.7, Def.11). On the other hand, there may well be theories, such as the theory of synthetic differential geometry ([KK2]), where interpretation into an arbitrary topos would either be impossible or would not make much sense. To resolve this indeterminacy is, of course, one of the general tasks of logical and mathematical research. In my view, this research is sufficiently advanced to claim that whenever the concept of 'space' is embodied in a type theory in a fundamental way then interpretation should be into some topos of spaces. By a 'fundamental way', I mean deictic concepts are involved. By a 'topos of spaces', I mean a topos in which geometric elements cohere. By 'geometric elements', I mean generic elements parameterised by some geometric object such as a 'point' or 'path'. By 'cohere', I mean that the restriction maps of the topos bring together the geometric elements into a recognisable geometric object. Thus a topos of spaces is a set-like category in which the objects are 'spaces', and what I have said gives some indication

that such a topos has variable sets as objects with the stages of definition indexed by geometric objects, which (by Yoneda) have geometric elements. To be sure, I do not mean that every type theory of a 'spatial' character can be interpreted into every topos of spaces, merely that it should be interpreted into such a topos. Thus the idea of a topos of spaces is to limit the possible candidates for a 'universe of discourse' in which to interpret type theories of a 'spatial' character. In a sense, the idea of a topos of spaces is to encode our knowledge of what, in general, a category of spaces is.

To clinch my first objective, that topoi of spaces encode knowledge arising directly out of the needs of Science, I shall argue that :

- a) the study of the covariance of observable space-time events requires cartesian closure and a covariant approach to variable sets;
- b) Lawvere's axiomatisation of a class of topoi known as the gros topoi of spaces encodes our most general idea of space as objects for describing the result of motion;
- c) the gros topoi of spaces include the familiar examples of spaces used in Science; and that
- d) the idea of topoi of spaces is a concept which includes both the gros topoi and any topos parameterised by a space (in a gros topos) such that objects have geometric figures in them. An important class of such topoi are the petit topoi of spaces, which are useful for studying many processes of interest to Science.

All our reasoning is mediated symbolically through language. I shall call the pre-symbolic influx of information into the processes of reasoning the iconic moment. I purloin this (Peircian) term from Ricoeur, who argues ([RI] p.189 & 199) that

"... the essential role of the icon is to contain an internal duality that at the same time is overcome ... the iconic moment involves a verbal aspect, in that it constitutes the grasping of identity within differences and in spite of differences, but based on a preconceptual pattern. Aristotelian seeing - 'to see the similar' - does not appear to be different from the iconic moment ... to grasp the relatedness of terms that are apart, is to set before the eyes".

I sloganise somewhat when I say the iconic moment relates 'seeing as' and 'saying as' and underlies the act of interpretation.

Nevertheless, to reason about the spatial arrangement of objects requires geometric knowledge. The iconic moment in such reasoning requires the use of the geometric diagram. The material substratum of a diagram shares all the relevant features of the objects designated in the pattern, so that we can detect a relation of resemblance. A diagram denotes the pattern of the objects by virtue of the work of resemblance in relating the diagram to the studied objects. A geometric diagram refers only to those features which relate to the spatial organisation or rearrangement of objects. Insofar as physics conducts experiments on non-intentional objects in space and time, the geometric diagram is well-suited iconically to the study of space-time events where theory can be investigated geometrically.

The work of resemblance postulates the importance of geometric figures, such as 'points' to reference location and 'paths' to describe motion. Resemblance would be impossible unless the 'points' and 'paths' cohered in a sensible fashion. Indeed the pre-symbolic influx of information into our reasoning processes is

grasped through the cohesion of the geometric figures. To be sure, 'points' and 'paths' are not the only geometric figures present, but they are the most obvious notions to effect the work of resemblance. Furthermore, there may be other less obvious notions at work such as 'infinitesimal', 'orthogonality', and 'dimension'. But in my view, it is sufficient for a general notion of 'space', if the reader rests content with the idea of Science needing models in which geometric figures cohere.

Habermas' argument, that the theories of a technical cognitive interest are law-like hypotheses which can be "... interpreted as statements about the covariance of observable events ..." ([HA1] p.308), sets up a requirement for geometry and mathematics to supply mathematical objects to study that covariation. The work of resemblance must match the covariation of observable space-time events with geometric objects which permit the rational study of covariation. Thus a Science with a technical cognitive interest needs a mathematics geared to this covariant approach of relating formalised statements of observable events to concepts of a spatial or geometric character, which facilitate their explanation. To study the covariation of events, I have argued that 'points' and 'paths' are the basic geometric figures. A 'point' references location by virtue of characterising a 'place' which is indivisible (= indecomposable) as it has no parts (= subobjects). A 'path' references an observable (smooth) change in location. To be sure, the concept of motion as the presence of a body in one place at one time and in another place at another time only describes the result of motion, and does not contain an explanation of motion itself ([LWC] p.136). However, treating 'points' as locations on a 'path' seems to be an elementary

prerequisite for the description of the covariation of space-time events. To move beyond description to an explanation of motion in terms of a technical cognitive interest requires (as Newton saw) a characterisation of motion which correctly expresses the continuity of time and space as the presence of the same body in two places at the same time ([GA2] p.13, [LWC] p.136). This is only an irreconcilable contradiction in a classical type theory when (as Lawvere puts it ([LWC] p.136)) "... we ignore the metaphysical opposition between points and neighbourhoods (introduced by the Platonic deification of points and revived by set theory) ...".

Such an opposition between 'points' and 'neighbourhoods' is briefly touched upon in Atkin's attempt to construct a covariant approach to physical observations. He notes that momentum "... is only well-defined if we can somehow attribute velocity to a vertex (= point) only" ([AT1] pp.202-3), and yet velocity (in cohomological terms) is a map from infinitesimal paths to quantities. Thus we have a seeming paradox that velocity is both defined on paths and points. In this light, the "... successful development of Newtonian physics and the parallel development of the mathematical tool - the differential calculus - were not merely coincidental" ([AT1] p.203). The development of 'infinitesimals' as the geometric figures which permit the study of motion and momentum as the presence of a body in two places at the same time is the "... burden of inventing the differential calculus" ([AT1] p.203). Thus whether the purpose is description or explanation, Science sets up a requirement for spaces as mathematical objects in which geometric figures cohere. To be sure, these remarks apply equally to metaphorical redescriptions of objectified or quasi-causal processes.

Unfortunately, this covariant approach to the mathematical objects required for Science has not always taken this rational form. Lawvere notes ([LWB] p.1) that :

"The mathematical background for theories of geometry, analysis, and continuum physics is usually considered to be the category of topological spaces (Sp) or the category of Banach manifolds (Man), with of course an infinite grade of smoothness conditions needed (apparently) for various technical theorems".

Lawvere calls the use of these categories the contravariant approach. In fact neither of these categories are topoi. Indeed, it has long been a source of embarrassment that Sp is not cartesian closed ([JT2] p.237). The situation with Man is even worse. It is not even complete; it lacks certain quotients, pullbacks, and coproducts ([D01] p.66). Yet the use of this contravariant approach has not always been dominant. Lawvere argues that ([LWB] p.1)

"... two centuries ago, many problems in the calculus of variations were correctly solved by mathematicians who, rather than defining a notion of 'open subset' for their function spaces, took the notion of 'path' as basic. Recognising the great importance of contravariant concepts such as open set (or real function) does not commit us to take these as the defining structure of a notion of space-in-general; they can be derived concepts in a theory where the covariant concept of geometric figures of some basic types, such as path, tangent vector, etc. are taken as primitive ...".

Essentially, Lawvere is urging a return to the covariant approach, in which variable sets parameterised by geometric figures have the topos-theoretic advantage of permitting a natural interpretation of physical theories. These natural advantages include the ease with which geometrical and physically-motivated constructions and axioms, such as function spaces (with good properties), can be simply effected - advantages denied to the contravariant approach, which tends to obscure the simplicity of these constructions.

This absence of well-behaved function spaces in Sp and Man is a considerable disadvantage, as a Science with a technical

cognitive interest has a crucial need for cartesian closure in obviously scientifically-motivated constructions. For example, in order to deal decisively with 'motion' or 'change', we need to treat the notion of 'path' or 'movement in the state-space'. So let \underline{E} be a cartesian closed category of smooth spaces and smooth morphisms, and let B be a space in \underline{E} representing a certain body, such as a system of 0-dimensional particles or a 3-dimensional fluid or solid body. Let T be a standard 1-dimensional space used to measure time, and let M be the ordinary flat 3-dimensional space. Proceeding naively, we act as if we were working with sets ([LWP]). A motion of B in M can be represented by an arrow $q: B \times T \rightarrow M$, which can be thought of as assigning to each particle in the body at each time instant the corresponding place in M . In the Galilean example, the motion q would be

$$\langle 1, t \rangle \mapsto 16t^2,$$

where the body B is idealised as a single 'particle'. The fact that \underline{E} has to be cartesian closed means that we can construct the space M^T , which might be thought of as the space of paths in M . The space M^B then parameterises all possible (and some impossible) placements of the body B in the space M , and we can usually obtain a functional $M^B \rightarrow M$ called the centre of mass. However, cartesian closure means that we have a diagram :

$$\begin{array}{ccc} & M^T \times T & \\ \hat{q} \times \text{id}_T \uparrow & \searrow \text{ev} & \\ B \times T & \xrightarrow{q} & M \end{array}$$

so that the exponential adjoint $\hat{q}: B \rightarrow M^T$ (obtained through lambda conversion) assigns to each particle in the body its path in M . In the Galilean example, \hat{q} is defined by the rule

$$1 \mapsto [t \mapsto 16t^2].$$

Indeed motion must be expressed in this way if we wish to compose it with the differentiation operator $M^T \xrightarrow{(\cdot)^0} V^T$, (where V is a vector space of translations) to obtain the velocity $v: B \times T \rightarrow V$. We have the diagram :

$$\begin{array}{ccc} & V^T \times T & \\ \hat{v} \times \text{id}_T \uparrow & & \searrow \text{ev} \\ B \times T & \xrightarrow{v} & V \end{array}$$

and if $\hat{v} = (\cdot)^0 \circ \hat{q}$, then by inverse lambda conversion $v: B \times T \rightarrow V$ expresses the velocity of each particle in the body B at any time instant. Galileo's falling stone would yield \hat{v} as

$$1 \mapsto [t \mapsto 32t]$$

and v as

$$\langle 1, t \rangle \mapsto 32t.$$

Furthermore, any coherent science with a technical cognitive interest will need to consider the motion in a third way : as a map $\hat{q}: T \rightarrow M^B$, which expresses the time dependence of the placement of the body in space. We need a diagram

$$\begin{array}{ccc} M^B \times B & \xrightarrow{\text{ev}} & M \\ \hat{q} \times \text{id}_B \uparrow & & \uparrow q \\ T \times B & \xrightarrow{\text{tw}} & B \times T \end{array}$$

in which the twist map $\text{tw}: T \times B \rightarrow B \times T$ is an isomorphism.

The map \hat{q} is then just the lambda transform of $q \circ \text{tw}$. Again,

(thinking of Galileo's falling stone, we have the rule

$$t \mapsto [1 \mapsto 16t^2]$$

as defining \hat{q} .

In all of these examples, the elementary nature of exponential adjointness stresses the need for cartesian closure at a simple conceptual level, even before any detailed consideration of what the 'smooth' nature of \underline{E} might be. Indeed, if the category of spaces, \underline{E} , is cartesian closed then unambiguous function space constructions exist with the good properties not only for empirical description but for explanation in terms of a technical cognitive interest. Thus I claim that Science needs categories of spaces which are (at least) cartesian closed. If we take a covariant approach and use variable sets which permit the unambiguous use of geometric figures, then our cartesian closed categories will be topoi too, suitable for the interpretation of scientific theories.

The definition of a topos of spaces as a topos in which geometric elements cohere may be just good enough for a good-reason-assay at the philosophical level, but from a mathematical perspective it is somewhat unsatisfactory. My argument seems to imply that the conceptual basis for topoi of spaces is rooted in mathematical experience with variable sets. But are all variable sets topoi of spaces, and if not, how are we to distinguish those which can be counted as categories of spaces from those which are not? In fact no complete mathematical definition of topoi of spaces has yet been achieved. However, it is possible to clarify mathematically certain properties a topos of spaces should have. Even if we cannot, as yet, give a full mathematical account of topoi of spaces as our idea of spaces-in-general, it is possible to produce models of (attempts to formalise) the idea. The principal starting point for any investigation of topoi of spaces is Lawvere's axiomatisation of the gros topoi of spaces ($[LWG, LWQ]$),

where the axioms select criteria for defining those topoi, which can serve the description of motion as 'change in location'.

Definition 1 A gros topos of spaces is an elementary topos, \underline{E} , defined over another topos, \underline{S} , with an essential geometric morphism $f: \underline{E} \rightarrow \underline{S}$ satisfying the following three axioms :

G1) $f_*: \underline{E} \rightarrow \underline{S}$ has a right adjoint $f^!: \underline{S} \rightarrow \underline{E}$; since f is essential, we have a string of adjunctions

$$f_! \dashv f^* \dashv f_* \dashv f^!;$$

G2) $f_!: \underline{E} \rightarrow \underline{S}$ preserves finite products, that is

$$f_!(1) = 1, \text{ and}$$

$$f_!(X \times Y) \cong f_!(X) \times f_!(Y);$$

G3) $f_!: \underline{E} \rightarrow \underline{S}$ sends the truth object Ω to 1, that is

$$f_!(\Omega) = 1.$$

In current scientific practice, \underline{S} is usually Boolean and \underline{E} is a Grothendieck topos. For my purposes it is convenient if \underline{S} is the honest-to-God category Set, then we can restrict our discussion to the geometric morphism $(\Delta, \gamma): \underline{E} \rightarrow \underline{\text{Set}}$, where \underline{E} is a variable set such as a presheaf category. To be sure axioms G1-G3 are formulated in the most general way possible to encourage the learning and development of mathematics. However, by restricting discussion to the case when \underline{S} is Set and \underline{E} is a (pre)sheaf category then matters are somewhat simplified and readily understood. Firstly, (Δ, γ) is essential when Δ preserves exponentials (Chap.7, Prop.21). For presheaf categories, Δ is full and faithful, thus $\underline{\text{Set}}^{\underline{C}^{\text{op}}}$ is always locally connected. When \underline{E} is a category of j -sheaves, then it is obvious that the topology will be for some site in which the constant presheaves are sheaves, since the constant presheaves are discrete objects ([BP] Theorem 16). We have the components functor $\prod_0: \underline{E} \rightarrow \underline{\text{Set}}$ as the left adjoint to

Δ (Chap.7, Prop.21). In the second place, γ only has a right adjoint if 1 is a representable ([LWQ] p.272, [JM] p.283). This right adjoint is known as the codiscrete functor $\nabla : \underline{\text{Set}} \rightarrow \underline{E}$. It sends a set X to a codiscrete object; that is, at stage A we have for $X(A)$ the set $\text{hom}(\gamma(A), X)$ with the obvious action on functions. If 1 is a representable then \underline{E} will need to be equivalent to a category of sheaves for some standard topology (Chap.7, Prop.17). Certain consequences flow from 1 being a representable.

Proposition 1 Each representable functor h_c in a presheaf category $\underline{\text{Set}}^{\mathcal{C}^{op}}$ is indecomposable and projective.

Proof : ([LS] Prop.9.2, p.171).

Thus a presheaf category satisfying axiom G1 is a model of intuitionist set theory with no empty types (Chap.7 Def.11).

(Clearly this rules out Graph since 1 is not representable). When \underline{E} is bivalent, that is $\underline{E}(1, \Omega) \cong 2$, then \underline{E} satisfies the stronger condition of being hyperconnected, since γ preserves Ω (Chap.7 Def.16). Thus with variable sets satisfying G1, there is a string of adjunctions

$$\prod_o \dashv \Delta \dashv \gamma \dashv \nabla \quad (\text{or verbally,} \\ \text{components} \dashv \text{discrete} \dashv \text{points} \dashv \text{codiscrete}).$$

To deal with axiom G2, we need the following.

Definition 2 A non-empty category \underline{A} is called filtered if

a) for every pair of objects A, A' there is a diagram

$$A \rightarrow A'' \leftarrow A';$$

b) for every pair of parallel arrows $f, g: A \rightarrow A'$ there is an arrow $h: A' \rightarrow A''$ such that $hf = hg$.

\underline{A} is cofiltered if \underline{A}^{op} is filtered.

Proposition 2 If \underline{C} is filtered (dually cofiltered), then

$\prod_{\mathcal{C}} : \underline{\text{Set}}^{\underline{\mathcal{C}}} \longrightarrow \underline{\text{Set}}$ (dually $\prod_{\mathcal{C}^{\text{op}}} : \underline{\text{Set}}^{\underline{\mathcal{C}^{\text{op}}}} \longrightarrow \underline{\text{Set}}$) preserves binary products.

Proof : ([JT1] Prop.2.57).

In fact the components functor fulfills the stronger condition of being left exact, when $\underline{\mathcal{C}}$ is filtered (dually cofiltered) ([JT1] Theorem 2.58). There is no general condition for a topos to satisfy axiom G3. However, as will become clearer, axiom G3 interacts in a strong way with axiom G2 to rule out Boolean topoi.

The motivation for these axioms is fairly simple : they select the simplest and most basic properties a set-like category of spaces-in-general should have. A generalised space as an object of $\underline{\text{Top}}$, fulfilling the requirements of axiom G1, is rather like the subcategory of locally connected spaces in $\underline{\text{Sp}}$. Recall that the forgetful functor $U : \underline{\text{Sp}} \longrightarrow \underline{\text{Set}}$, which sends a topological space to its (carrier) set of points (yes, points are continuous maps from the one-point space), has both left and right adjoints : the 'inclusions' of spaces with the discrete and indiscrete (= codiscrete) topology ([D01] p.36). It is well-known that, in general, the 'inclusion' of discrete spaces does not have a left adjoint. However, such a left adjoint does exist for the full subcategory of locally-connected spaces ([ML] p.131). Thus axiom G1 mirrors an important class of non-pathological topological spaces in terms familiar to those taking the contravariant approach. Axiom G2 marks the passage to qualitative considerations of path-connectedness and homotopy, which are necessary for a covariant approach in Science. It is well-known that path-connectedness implies connectedness but not conversely ([WA] Theorem 3.4). Thus any axiom relating to (local) path-connectedness is a stronger condition independent of considerations of (local) connectedness.

The components functor can certainly be construed as an analogue of the path components of geometric topology. It results from coequalising the domain and codomain maps (Chap.7 Prop.21) of a discrete fibration (= small category); thus it identifies those objects of a category which are connected along a path of composable forward or backward arrows ([BW] p.10). But to make the analogy complete, axiom G2 insists that the components functor preserves binary products, since a product space $X \times Y$ implies that $\prod_o(X \times Y) = \prod_o(X) \times \prod_o(Y)$ ([MS] Theorem 7.1). We interpret $\prod_o(X)$ as the number of path components of X , and when $\prod_o(X) = 1$ we say X is (path) connected. Forgetting the topos-theoretic aspects for a moment, axioms G1 and G2 together mirror certain minimal requirements a category of path and locally connected 'spaces' should have. However, axiom G3 marks the crucial move to a topos (and a cartesian closed category) rather than a category of spaces. The truth object Ω of \underline{E} is a Heyting algebra object in \underline{E} , which in turn means that Ω has the structure of a join semilattice (or monoid with zero). In the presence of axiom G2, its being connected (G3) implies that Ω is contractible, in that $\prod_o(\Omega^X) = 1$ for all X in \underline{E} ([LWG] p.181). Thus every space X has an embedding into an injective and contractible space through the singleton map $\{ \}: X \longrightarrow \Omega^X$ ([LWG] p.181). Taken together axioms G2 and G3 imply that \underline{E} cannot be a Boolean topos. Since \prod_o has a right adjoint, it preserves all colimits (Chap.5 Prop.3), so that $\prod_o(1 + 1) = \prod_o(1) + \prod_o(1) \neq 1$, and $1 + 1 \cong \Omega$ in a Boolean topos. Note that this does not say that Ω cannot be contractible in a Boolean topos. For example, the Boolean algebra

$$\Omega_{\neg\neg} = \boxed{\begin{array}{ccc} & \tau & \\ \perp \bullet & \xrightarrow{\quad} & \bullet \top \\ & \xleftarrow{\quad} & \\ & s & \end{array}}$$

for the double negation topology on RGraph is contractible, as can be seen. What it does say is that one cannot have the components functor preserving path and local connectedness and have a contractible truth object at the same time in a Boolean topos. In practice, variable sets with a non-Boolean truth object are the norm in algebraic and differential geometry ([LWV] p.106), supporting Lawvere's contention that "... no logic stronger than intuitionistic logic can be valid for sets varying in any serious way ..." ([LWV] p.105). Thus in the passage from categories to (gros) topoi of spaces, axiom G3 forces the logic of these topoi to be non-classical.

The gros topoi provide the basic notions of spaces-in-general for a Science (with a technical cognitive interest) needing to describe motion as a change in location. Axiom G1 captures the notion of locally connected spaces with indecomposable and projective points to reference location. Whatever geometric figures cohere in E, then axiom G2 makes them cohere in path-connected components just good enough to reference change in location. Finally axiom G3 means that our topoi of spaces will be models of non-classical type theory, and enable possibilities which would only seem contradictory in the classical case.

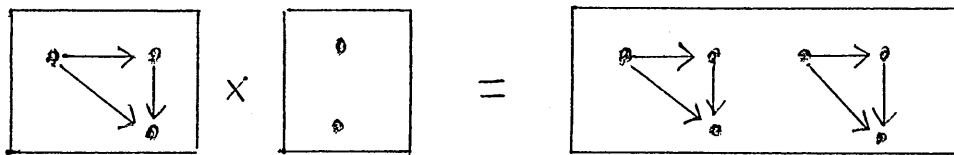
Perhaps the simplest category of a space-in-general is RGraph. Here points (= nodes with degenerate loops) reference location, and edges reference possible movements between points. Thus RGraph is the most primitive candidate for a description of motion as a change in location.

Proposition 3 RGraph is a gros topos of spaces.

Proof : Lawvere ([LWG]).

With RGraph we set S to Set. Of course (Δ, δ) is essential, and we have the components functor \prod_0 as a left adjoint to Δ . Since 1 is a representable, then 1 is indecomposable and projective, and δ is left exact (Chap.7 Def.11) and the inverse image of a geometric morphism whose right adjoint is the codiscrete functor $\nabla : \underline{\text{Set}} \rightarrow \underline{\text{RGraph}}$. It sends a set X to the codiscrete graph, where the points are the elements of X and there is exactly one edge in each direction between every pair of points. Thus RGraph satisfies G1. The components functor \prod_0 is just the connected components of a reflexive graph in the usual sense (Chap.7, Prop.21 ff.). It is easily seen that 1 is connected ($\prod_0(1) = 1$). Indeed connectivity is preserved in products.

For example, the product



is preserved under the action of components. It is easily seen that $\underline{\text{Set}}_{\underline{M}(2)^{op}} \cong \underline{\text{Set}}_{\underline{\Delta}_1^{op}} \cong \underline{\text{RGraph}}$ satisfies axiom G2 since M(2) is cofiltered, with every arrow from I to 1 as a retraction. Finally, it is obvious that the truth graph is connected and contractible, satisfying axiom G3.

The last Proposition suggests the following.

Proposition 4 Categories of generalised graphs $\underline{\text{Set}}_{\underline{M}(\tau)^{op}}$ for $T > 2$, and categories of simplicial sets $\underline{\text{Set}}_{\underline{\Delta}_D^{op}}$ for $D > 1$ are gros topoi of spaces.

Proof : ([LWQ]).

Indeed it is readily seen that a change in parameter, T in the case

of graphs and D in the case of simplicial sets, would make no difference to the proof of Proposition 3. Thus we have many examples of gros topoi, with fibred products and quotient constructions with exactness properties similar to those in the naive category of constant sets. Indeed, as Lawvere argues ([LWB] p.2) :

"The axiomatics at the category level is also valuable because there are many related categories which immediately come up. For example, if \underline{E} is a gros topos of spaces and G is a group in \underline{E} while B is an object of \underline{E} then the categories \underline{E}/B of B -parameterized families of spaces, \underline{E}^G of actions of G on spaces in \underline{E} , and \underline{E}^G/B (of central interest to bifurcation theory) are all categories which satisfy the same axioms as \underline{E} , as does (a reasonable determination of) the category of all objects of \underline{E} equipped with affine connection."

Thus, for those wishing to take the covariant approach, there are many topoi meeting the stringent requirements of axioms $G1-G3$, which distinguish the gros topoi from variable sets in general.

The availability of gros topoi of spaces for a covariant approach is not quite the same thing as saying that Science needs them, even if it is admitted that a covariant approach is suitable for Science. Much of current mathematical practice takes the contravariant approach using objects and morphisms drawn from \underline{Sp} and \underline{Man} . However, I claim that the non-pathological spaces and manifolds (and their morphisms) useful for Science can be included in some gros topos of spaces. If it is admitted from a logical point of view that it is better to be in a topos rather than in an arbitrary category, and if in practice physicists and engineers continue to argue naively as if they were in the category of sets, then I claim that the rational form of the notion of space-in-general (suitable for describing the motion of moving matter) is embodied in the concept of the gros topoi of spaces. Science needs the gros topoi, in that it needs always to find the rational form

for its conceptions.

The most useful objects of Sp for Science are the sequential spaces which permit notions of 'symbolic dynamics', where points reference states-of-affairs and sequences of points reference trajectories of objectified processes. As Isbell puts it "... sequential spaces seem to suffice for 'geometric topology' ..." ([IS] p.197). Here, categories like $\text{Set}^{\underline{M}(\tau)^{op}}$ come into their own, where an 'edge' at stage I has τ arrows to the points at stage 1. These 'edges' can represent 'paths' or 'trajectories'. The τ arrows send the 'edge' to the t 'th point on the 'path'. It is tempting to set τ to the cardinality of an infinite but countable set, to obtain the category $\text{Set}^{\underline{M}(\omega)^{op}}$. However, $\text{Set}^{\underline{M}(\omega)^{op}}$ is 'too large', for clearly such variable sets may include sequences of elements which are not sequences in the topological sense. Nevertheless, Johnstone shows ([JT2]) that, by the simple expedient of endowing $\text{Set}^{\underline{M}(\omega)^{op}}$ with the canonical topology, one can obtain a topological topos, denoted by Tsp, in which a topological space can be associated with a sheaf in a subtopos of $\text{Set}^{\underline{M}(\omega)^{op}}$. The generic figure 1 is the one-point space. The set of points of the generic figure I (modulo the canonical topology) is the set $N^+ = N \cup \{\infty\}$, where N is the set of natural numbers. N^+ is endowed with the topology of one-point compactification of the discrete space of natural numbers. Thus, (the Karoubi envelope of) $\underline{M}(\omega)$ can be embedded (by Yoneda) in Tsp to obtain the full subcategory of Sp consisting of N^+ and 1. Indeed, the monoid of endomorphisms on N^+ is just the action of $\underline{M}(\omega)$. It is now possible to define a functor $H: \underline{Sp} \rightarrow \underline{Tsp}$, which sends a topological space X to its sets of points at stage 1 and to $\underline{Sp}(N^+, X)$ at stage I. Thus the set

of convergent sequences in X are regarded as figures at stage I , and the points of X are the generic figures at stage 1. The T arrows from $X(I)$ to $X(1)$ identify the t 'th point of the convergent sequence, and the single arrow from $X(1)$ to $X(I)$ picks out the constant sequence equal to the t 'th term. From a covariant point of view, the cohesion of these spaces comes from identifying the t 'th point of a generalised edge called a 'sequence of points' by topologists.

Proposition 5 The functor $H: \underline{Sp} \rightarrow \underline{Tsp}$ is faithful; and it is full when restricted to the subcategory of sequential spaces.

Proof : ([JT2] Lemma 2.1).

I do not have the space here to give a full survey of Johnstone's results. However, \underline{Tsp} contains all the first-countable spaces, and hence metric spaces. Since the sequential property is inherited by quotients, then every quotient of a metric space is sequential and thus an object of \underline{Tsp} ([JT2] pp.240-1). However, unlike the categories of topological and sequential spaces, \underline{Tsp} has all the good properties of being a (gros) topos. For an explicit account of the site of definition and the bivalent nature of $\underline{\Omega}$, the reader is referred to the lengthy details ([JT2]) needed to fully comprehend the construction of this topos. However, two points are worthy of attention. Firstly, one of the attractions of this topos is that the Dedekind real number object is just the reals with the usual topology ([JT2] Prop.4.4). Thus all those constructions in topology requiring the reals with the usual topology may find their rational form in \underline{Tsp} . Secondly, an even deeper result is :

Proposition 6 There is a geometric morphism, actually a surjection, $r: \underline{Tsp} \rightarrow \underline{Set}^{\Delta^{op}}$, such that

- a) if K is a simplicial set, $r^*(K)$ is the geometric realisation of K (considered as a sequential space and hence as an object of

Tsp); and

b) if X is a sequential space, $r_*(X)$ is the singular complex of X .

Proof : ([JT2] Theorem 8.1).

It seems to me that one of the subterranean ideas in algebraic topology is that singular complex and geometric realisation are adjoint functors. With Proposition 6 they are actually direct and inverse images of a geometric morphism between two gros topoi of spaces. Thus spatial notions which are usually developed contravariantly are easily construed as objects and arrows in a gros topos of spaces, with the advantage that we can talk (and refer covariantly to objects) with the internal language of Tsp.

The construction of the topological topos provides the basic insight into the construction of another interesting gros topos due to Lawvere. He was interested in finding a covariant definition of topological spaces suitable for studying the foundational problems of continuum mechanics. He argued that the important feature of a space is the set of continuous paths which can be followed in it ([LWB]). He therefore sought to establish the notion of 'path' as a generic figure. Lawvere's topos (denoted by Law) is rather like Tsp, in that the canonical topology is placed on the generalised graphs of $\text{Set}^{M(\tau)^{op}}$, where T is now the cardinality of the continuum. Thus the figure I might now be thought of as the unit interval $[0,1]$, with a monoid of continuous endomorphisms on it, such that the points $1 \dashrightarrow I$ are the points of $[0,1]$ in the usual sense. The figure 1 is, of course, the one-point space. Thus, for any space X , the set $\text{Path}(X)$ of continuous paths is a right M -set for which all the M -sets are sheaves. Essentially, we have a faithful functor $\text{Sp} \dashrightarrow \text{Law}$ ($X \dashrightarrow \text{Path}(X)$), which "... is full

on a reasonably large subcategory of Sp (which includes, for example, all the C-W Complexes and all the topological manifolds) ..." ([JT2] p.239). In Law, the maps $I \times I \dashrightarrow I$ are just the usual continuous functions of two variables ([LWQ] pp.275-6),

"... which is mildly surprising, but not too difficult to prove. The corresponding statement for the monoid of smooth ($= C^\infty$) self maps of the line is surprising (once one realizes that one has to show that all the higher formally defined partial derivatives are the actual partial derivatives so in particular commute) and rather difficult to prove (see [LWB],[FK]). Having calculated Y^I , even more interest attaches to the natural path functionals $Y^I \dashrightarrow Y$. In both of the examples mentioned, the real line R determines an object ... with just the reals as points by defining its elements ... (at stage) ... I to be just the continuous (resp. smooth) paths in R . Since multiplication is continuous it gives rise to a morphism $R \times R \dashrightarrow R$... and hence to a left action $R \times R^I$ by $(a.f)(x) = a(f(x))$. Thus (in the smooth case) one can look for the object of linear functionals

$$\text{Lin}_R(R^R, R) \dashrightarrow R^{R^R},$$

whose points are just the morphisms $\gamma : R^R \dashrightarrow R$...".

Nowhere is the connection between a gros topos as variable sets and a covariant approach in Science, requiring just such paradigmatic figures as 'points' and 'paths', as explicit as this.

The gros topoi, discussed so far, have one thing in common: they are graphs or subtopoi of generalised graphs, in which the topology is standard in order to preserve discrete objects. The representables form a full subcategory (of two objects) of the spaces being modelled. All this suggests taking some full subcategory \underline{C} (including 1) closed under the formation of subspaces of some concrete category of spaces \underline{Z} , and then with the aid of some standard topology studying the resulting category of j -sheaves. There is always a faithful functor $H: \underline{Z} \dashrightarrow \underline{E}^j$ which is full when restricted to \underline{C} , and generally full on a much larger subcategory of \underline{Z} . Obviously, this programme can have its difficulties : a site of definition could be large and unwieldy,

and it might be difficult, in practice, to compute anything directly about it. Nevertheless, this is a convenient route for the study of differential geometry. Kock's book ([KK2]) develops the theory of synthetic differential geometry from a topos-theoretic viewpoint, which is just differential geometry in terms of set-like reasoning. This is a vast subject and cannot be entered into in any detail here. Nevertheless, I wish to draw the reader's attention to three elementary aspects of models of this theory. Firstly, that models of the theory are variable sets also satisfying the axioms for a gros topos. Secondly, that the usual categories of manifolds (used extensively in the Sciences) can be embedded in such a topos. And thirdly, that these variable sets possess a distinguished object (= space) of infinitesimals suitable for reasoning about differential calculus in a non-classical fashion.

Reyes seeks models of the theory in variable sets of the form $\text{Set}^{\underline{\mathcal{C}}^{\circ p}}$, where $\underline{\mathcal{C}}$ is a category of (small) rings ([RY]). The objects of $\underline{\mathcal{C}}$ are quotient rings of the form $C^\infty(R^n)/I$, where $C^\infty(R^n)$ is the ring of smooth functions $R^n \rightarrow R$, and I is a germ-determined ideal. That is, $f: R^n \rightarrow R$ is an element of an ideal I , such that for all $p \in R^n$ there exists an open neighbourhood V_p around p with functions g and $h \in C^\infty(R^n)$, where $f = g + h$, $g \in I$, and h vanishing on V_p ([KK2] pp.230-1). Intuitively we think of the pair (n, I) as the locus of the set of equations $f(p) = 0$ for $f \in I$. If $A = C^\infty(R^n)/I$ and $B = C^\infty(R^m)/J$ are rings in $\underline{\mathcal{C}}$, a morphism $B \rightarrow A$ in $\underline{\mathcal{C}}$ is the dual of a ring homomorphism $A \rightarrow B$ which is induced by composition with a smooth function $R^n \rightarrow R^m$. Explicitly, a morphism $[\phi]: B \rightarrow A$ in $\underline{\mathcal{C}}$ is an equivalence class of smooth functions $\phi: R^n \rightarrow R^m$, such

that for all smooth $f: \mathbb{R}^m \rightarrow \mathbb{R}$, $f \in J$ implies $f \cdot \phi \in I$,

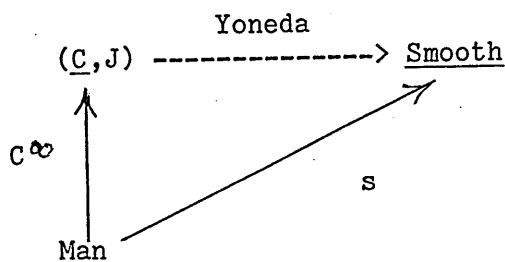
$$\begin{array}{ccc}
 \mathbb{R}^n & \xrightarrow{\phi} & \mathbb{R}^m \\
 \searrow f \cdot \phi & & \swarrow f \\
 & \mathbb{R} &
 \end{array}$$

Any two such functions ϕ and ϕ' are equivalent if for all projections $p_i: \mathbb{R}^m \rightarrow \mathbb{R}$ ($i = 1, \dots, m$) then $p_i \cdot \phi - p_i \cdot \phi' \in I$. In fact \underline{C} contains all isomorphic copies of rings of the form $C^\infty(U)/I$ where U is an open subspace of some Euclidean space \mathbb{R}^n and I is a germ-determined ideal. To be sure, a full understanding of the construction of \underline{C} depends on a good grounding in commutative algebra and algebraic geometry. However, for the purposes of this thesis all that is essential is an intuition into the idea that a category \underline{C} of small rings can be founded as the domain of contravariant Set-valued functors. Strictly speaking, these presheaves are not altogether a good model for reasoning about differential geometry. Actually, a standard (but not canonical) topology (known as the open cover topology) is required for a range of technical reasons which are beyond the scope of these notes ([KK2] Chap. 3.7). I shall denote the open cover topology (in the sense of Lawvere-Tierney) by j , the topos $\underline{\text{Set}}^{\underline{C}^{\text{op}}}_j$ by Smooth, and the site of definition by (\underline{C}, J) , where J is the Grothendieck topology associated with the open cover topology.

Now the topos Smooth is an example of a gros topos (see [KK2] pp.292-3). Reyes quite specifically gives details ([RY] p.76) of the discrete and codiscrete functors which are adjoint to the points functor (axiom G1). Proofs that Smooth satisfies axioms G2

and G3 require extensive development of algebraic and differential geometry which is beyond the scope of these notes. However, some pointers that Smooth does satisfy these axioms will be mentioned in what follows.

Now the site (\underline{C}, J) contains copies of all compact manifolds. Basically we work with the following diagram of functors:



in which the functor s sends a compact manifold to a j -sheaf in Smooth. The full and faithful functor $C^\infty : \underline{\text{Man}} \rightarrow (\underline{C}, J)$ sends a manifold M to the set of smooth functions $M \rightarrow R$, and denoted by $C^\infty(M)$ ([KK2] Chap. 3.5). Now the image of $C^\infty(M)$ in (\underline{C}, J) is a subcategory of finitely presented rings in (\underline{C}, J) , which on composition with Yoneda yields the full and faithful functor $s : \underline{\text{Man}} \rightarrow \underline{\text{Smooth}}$ ([KK2] pp.216-8). Not only is the functor s a full embedding, but it has the excellent qualities of preserving products, (transversal) pullbacks, partitions of unity, finite intersections of open subspaces, open covers, compactness, and connectedness ([KK2] Part3, [MR] p.65). In turn, the preservation of products and connectedness of manifolds when embedded in Smooth gives a pointer to the importance of axiom G2. In particular, in Smooth we have smooth spaces like:

$$\begin{aligned}
 1 &= s(R^0) && \text{the generic point,} \\
 L &= s(R) && \text{the smooth geometric line, and} \\
 [0,1] &= s([0,1]) && \text{the smooth unit interval.}
 \end{aligned}$$

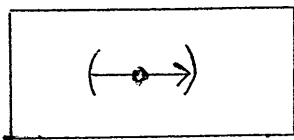
Of course, differential geometry in Smooth is not quite the same as

differential geometry classically conceived, for Smooth contains not only 'constant quantities' such as $0, 1, \pi, \sqrt{2}$ (defined at stage 1), but also 'variable quantities' (defined at other earlier stages) as well. Nevertheless, Smooth is 'well-adapted' to study the 'analytical' differential calculus for manifolds, for the real number object is the type of smooth geometric line. From this perspective, manifolds are objects in a topos and it is now possible to rationally reconstruct classical differential calculus in terms of set-theoretic reasoning ([KK2]).

For instance, Smooth has useful structure available that only degenerates in the classical context. For example, the subobject D of L defined (in the internal language) by

$$D = \{d \in L \mid d \cdot d = 0\}$$

is the type of first-order infinitesimals. This is somewhat different from the one-point space (which is the only nilpotent in the classical case), as can be illustrated by the fact that functions from D correspond to tangent vectors. For every manifold M there is a canonical isomorphism relating to the tangent bundle $T(M)$, that is $s(T(M)) \cong (s(M))^D$ ([RY] p.76). We can picture an element of $s(M)^D$ as



on a suggestion from Kock and Reyes ([KR] p.194), in which an infinitesimal is '... a certain definite (but small) piece of the line ...' associated to a point m in M . The first principles of differential geometry revolve around showing that any element (= function) in L^D is a straight line.

Proposition 7 Smooth satisfies

$$(\forall g \in L^D)(\exists ! b \in L)(\forall d \in D) g(d) = g(0) + d \cdot b$$

such that $L^D \cong L \times L$.

Proof : ([RY] p.78, [KK2] pp.3-6).

Of course, the Dedekind reals in Set fail to satisfy the formula of Proposition 7. But the line object L and the infinitesimals D in Smooth do meet these elementary requirements. Kock points out that Proposition 7 is ([KK2] p.6) :

"... incompatible with the law of the excluded middle. Either the one or the other has to leave the scene ... this means that the logic employed is 'constructive' or 'intuitionistic'. We prefer to think of it just as 'that reasoning which can be carried out in all sufficiently good cartesian closed categories'."

The incompatibility of the law of the excluded middle with Proposition 7 is a pointer to the significance of axiom G3. Thus this gros topos provides a universe of discourse to confound Bishop Berkeley.

The axioms of gros topoi seem to encompass all the familiar examples of 'spaces' known to Science. Thus the idea of space-in-general finds a rational form in such topoi as RGraph, Tsp, Law, and Smooth. All these topoi are models of axioms G1-G3. To be sure, these topoi differ in that some of them meet further axioms of relevance to Science. However, the important point is that the gros topoi englobe the study of geometrical objects. It would be possible to argue that the gros topoi have the features relevant to Science with a technical cognitive interest, and that a scientist could effect all his calculations in such a topos. However, this seems unduly restrictive, and the following examples provide grounds for exercising some caution here. Firstly, a localic topos cannot be a gros topos since it cannot satisfy axioms G2 and G3 at the same time ([LWG] p.182). Secondly, Graph fails to satisfy axioms G1 and G2 ([LWG] p.183). In the first case, localic topoi

(in their guise as sheaves on a topological space) have been used extensively in mathematics since 1945 ([GY]), and have been successfully transferred to modern physics. For example, the twistor theory of Penrose rests heavily on the theory of sheaves for operational reasons, since physical theories are never perfectly localised. Indeed, the logic of twistors is intuitionistic, usually determined by the Heyting algebra of open sets of the 2-sphere ([JZ]). In the second case, irreflexive graphs have been used successfully by mathematical sociologists. Thus, in practice, scientists can be found using topoi which do not meet the requirements for a gros topos.

Nevertheless, I claim that these scientists are using a topos of spaces, in the sense that these (non-gros) topoi have objects in them in which geometric elements cohere. Thus, in the case of the sociologists, irreflexive graphs have recognisable geometric elements such as nodes and edges; and in the case of the twistor theorists, precise global points are required as well as neighbourhoods for quantised events. Yet in all these cases, the scientists are using a topos parameterised by an object from a gros topos of spaces. For example, the twistor theorists are using a localic topos parameterised by a space (= 2-sphere) which may be considered as a Heyting algebra object in Law; and it can be shown (as below) that the sociologists are using a topos (Graph) parameterised by the 'generic' loop in RGraph. Thus we are lead to define a topos of spaces either as a gros topos or a topos parameterised by objects and morphisms in a gros topos.

The appearance of non-gros topoi of spaces should not be too surprising. Lawvere puts it in a Hegel-like fashion ([LWQ] p.261 &

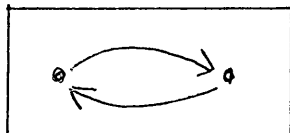
"... 'Being is doing', and hence particular being is known (at least partly) by what it can do. If B is an object in a 'gros' topos E of cohesive active sets, what it can do is to continuously parameterize and dynamically act on mathematical structures ... I have used the term 'space' as short for cohesive/active set; already in Grassmann it was clear that space is generated by, and lays the foundations for, motion and hence general spaces have aspects of both."

Definition 3 A petit topos of spaces is an elementary topos, $\underline{S}_{\underline{E}}(B)$, where $f: \underline{E} \rightarrow \underline{S}$ is a gros topos of spaces and B is an object of \underline{E} , such that $\underline{S}_{\underline{E}}(B)$ is the full subcategory of the slice topos \underline{E}/B , consisting of all objects E in \underline{E}/B such that the following diagram is a pullback :

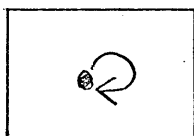
$$\begin{array}{ccc} f_*^* & f_*^* & E \xrightarrow{\quad} E \\ \downarrow & \downarrow & \downarrow \\ f_*^* & f_*^* & B \xrightarrow{\quad} B \end{array}$$

170

point of this mathematical manoeuvre is to ensure that the objects in $\underline{S}(B)$ have discrete fibres (yes, something 'active', like particle spin, is being forgotten). For example, if \underline{E} is RGraph and B is the special case



then the topos $\underline{Set}(B)$ is equivalent to the category of bipartite graphs. Generalising this to any B in RGraph, then the topos $\underline{Set}(B)$ might be thought of as categories of B-partite graphs ([LWG]). Given a morphism $g:B \dashrightarrow B'$ in RGraph, the topos $\underline{Set}(B)$ behave with excellent functorial comportment, in that objects of such topos can be transported 'continuously' from $\underline{Set}(B)$ to $\underline{Set}(B')$, and vice-versa. Indeed, (g^*, g_*) is an essential geometric morphism ([JT1] Theorem 2.34). Furthermore, it can be shown that if L is the 'generic' loop



then those L -labelled graphs (sending all non-degenerate loops and edges to the single non-degenerate loop of L) form the topos $\underline{Set}(L)$ (of L -partite graphs) which is equivalent to Graph ([LWG] p.185). Clearly under this definition, Graph is a simple example of a petit topos, in which the objects vary over the 'generic' loop.

We can conceive of categories $\underline{Petit}(\underline{E}, \underline{S})$, in which the objects are the petit topos $\underline{S}(B)$ where B is any object of a topos of spaces $\underline{E} \dashrightarrow \underline{S}$, and the morphisms are geometric morphisms. Clearly, under this definition $\underline{S}(1) \simeq \underline{S}$, just as $\underline{E}/1 \simeq \underline{E}$. If \underline{E} is RGraph then the topos of 1-partite graphs (or points), $\underline{Set}(1)$, is equivalent to Set. This makes Set an example of a generalised space. However, it is instructive to invert the usual way of

looking at things, and consider Set as derived from some E. When the scientist passes from studying his usual variable sets to some constant sets he is usually freezing some variation. As Lawvere puts it ([LWC] p.136) :

"Every notion of constancy is relative, being derived perceptually or conceptually as a limiting case of variation and the undisputed value of such notions in clarifying variation is always limited by that origin. This applies in particular to the notion of constant set, and explains why so much of naive set theory carries over in some form into the theory of variable sets."

Clearly no harm can come through using Set (in the inversion of the usual way of looking at things) in which we pass from variable sets to constant sets and back again. However, we might not wish to think of Set as a generalised space in isolation from its role in Petit(E,Set). If Set is construed in splendid isolation from this framework then there are obvious dangers in forgetting the 'geometry' underlying the 'points'. One way of looking at the 19th Century development of Analysis is to (anachronistically) think of it as living in Set Smooth (1).

The class of petit topoi is one (rather important) way to associate an object of a gros topos with a topos to obtain a topos of spaces. It is perhaps a paradigm of future attempts to broaden the concept of a topos of spaces, by loosening in a disciplined fashion the stringent requirements of a gros topos. However, I feel that enough has been done to illustrate that we can produce models of (attempts to formalise) the notions of a topos of spaces, and have good ideas about their prospective developments. In particular, the claim made in Chapter 4 that the conceptions of space required by the Human Sciences are essentially no different from that required by the Natural Sciences has been fully substantiated. Whereas a natural or quasi-causal scientist will

emphasise the 'active' aspects of (metaphorically redescribed) motion, a research worker in the Human Sciences (with a good-reason-assay in mind) may only need to emphasise the 'cohesive' aspects of 'structure' and so a petit topos may be adequate for his needs. However, there can be no hard and fast rule here. The choice of the topos (of spaces) will depend on the model-theoretic job it will be required to fill.

Few scientists these days will concur with the savage Duhemian rejection of models ([HE2]). One cannot remain frozen in the Formal. The Conceptual is needed also, even though most scientists "... do not regard models as literal descriptions of nature but as standing in a relation of analogy to nature" ([HE1] p.201). One understands the world through metaphorically redescribing it with the model. Of course, one can interpret a type theory into the topos generated by it. But this is constructed linguistically and is not necessarily a topos of spaces. An interpretation, properly conceived, is a logical functor from the topos generated by the theory into such a topos of spaces. Even the relationists (Chap.4) have some foothold in the Conceptual with their idea that space is a network or lattice of relations. Of course, these relational notions are notoriously unclear. Perhaps we might persuade them that this network should take the form of a small category. In this case their network of relations could take the form of variable sets. If we were more insistent, we might persuade them that these variable sets were (gros) topoi of spaces, and that their abstract set of relations were the same as the 'concrete' spaces of the Newtonians. Ontological issues apart, the rational form of space is as a variable set in which geometric elements cohere.

If formalised fragments of scientific discourse are to be cast in terms of type theory, and if such theories elucidate relations between elements construed as geometric figures, and if theories must be interpreted in an elementary topos, then it might be thought unsurprising that interpretation must be in some topos of spaces, in that such topoi are categories in which geometric figures cohere. Nevertheless, there are examples of scientists attempting to understand notions involving spatial ideas with mathematical concepts which are NOT drawn from any topos of spaces. Perhaps this might be taken as empirical evidence refuting my first objective. I do not think this proposition can be seriously entertained. On the contrary, I believe that when we examine such examples, we are likely to witness troublesome mathematical and logical difficulties, and a struggle to realign the mathematical ideas in terms of some topos of spaces. From the perspective of my first objective, this is understandable. If topoi of spaces encode our knowledge of space, then there is a mathematical and logical compulsion to realign our mathematical ideas with them.

To support this argument, I shall outline two studies in the realignment of mathematics drawn from my own research interests. Perhaps my own experience is untypical. Nevertheless, I was somewhat surprised to find such support so readily to hand. Possibly the reader might find other examples drawn from their own research experiences. My two studies focus on :

- 1) Varela's use of a calculus of self-reference for studying the autonomy of autopoietic systems in the Biological Sciences; and
- 2) Atkin's use of some concepts of algebraic topology in the Social Sciences.

Both Varela and Atkin are scientists with rich imaginative ideas. I shall not attempt to engage in any detailed critique of their work, or to explore the plenitude of interesting philosophical, scientific, or mathematical ideas that may stem from any serious consideration of their work. This would take us too far afield. I shall deliberately restrict my reflections to questions regarding the realignment of their mathematical conceptions with topoi of spaces. It is no part of my argument that either Varela or Atkin (or their coworkers) were necessarily concious or self-reflective about the realignment of their conceptions with some topos of spaces. But it is a part of my argument that they are logically compelled to do so. On the other hand, if research groups are concious and self-reflective about the knowledge encoded in topoi of spaces, then this should lead to a rapid diagnosis of any mathematical difficulties caused by not being in such a topos.

CHAPTER NINE

VARELA AND SELF-REFERENCE

IN BIOLOGICAL SYSTEMS

Varela's requirement for a conceptual analysis of self-reference in biological systems stems from his studies of the lymphoid network. Contemporary interest in the immune system postulates the existence of a biochemical network resulting from the action by determinants of antibody molecules. Varela argues ([VA] p.221) that :

"It is misleading to view the activity of individual clones in isolation, because once the first (antigen-binding) antibodies are formed, they generate anti-antibodies (anti-idiotypic), which in their turn would generate anti-anti-antibodies, and the process would grow in an everbranching tree involving the whole lymphoid system. It is obvious, therefore, that the system exhibits closure, which modulates the magnitude of this process; in other words, at a certain point this cascade of stimulating reactions must start 'biting its own tail' and leave the system in a new state of equilibrium."

Thus Varela proposes that the immune system be regarded as an autonomous unit; that is, as a network of cellular interactions that at each moment determines its own identity. Cellular interactions give rise to the connectivity of the lymphoid network, which consists of the totality of the lymphoid tissue. The study of specific immune response against the huge variety of specific antigens is viewed as the response of a 'system-whole' to environmental perturbation. The 'organisation' of the system exhibits cumulative effects as an 'everbranching' process, which must be 'closed' to obtain a new state of equilibrium (or behaviour). Since the 'everbranching' process of causal effects is closed, then the 'system-whole' must 'bite its own tail'. The

latter is called self-reference or re-entry.

Some insight into Varela's perspective can be gained from Eigen's work on the hypercycle. Eigen studied the successive steps of stability in biochemical interactions through a system of non-linear differential equations derived from non-equilibrium thermodynamics. Selective pressures are brought to bear in the processes of molecular evolution. The circular concatenation of processes is called the hypercycle, and is postulated as the unit of selection in early life ([VA] p.28). Of course, Eigen's work is not equivalent to a formalisation of an autonomous system. Starting from a need to use the differentiable approach, Eigen concentrates on the network of interactions and their temporal invariances, and deliberately disregards the way in which these reactions constitute a unit in space. Nevertheless the hypercycle, as a biochemical process of 'biting its own tail', is an example of a self-referential process. It is to the logic of self-referential processes that Varela turned.

Varela thought that he had found a general mathematical and logical language in Spencer-Brown's Laws of Form ([SB]). He developed a calculus of self-reference as an extension to Brown's work, and thought that this could not only illuminate the philosophical language, which his biological studies required, but also act as a very general test bed for ideas about self-reference in biological systems. He never intended the calculus as a realistic representation of any biological process. Unfortunately Varela's calculus exhibited various logical difficulties. In joint work with Goguen, he redefined the calculus in terms of an object in a cartesian closed category ([VG]). Essentially, Goguen drew on

contemporary work in the theory of computation and fixed-point algebras in modal logic ([SY]), to realign Varela's notions with a special class of cartesian closed categories, in which an infinite sequence of algebraic processes (= biochemical interactions) could be identified with a fixed-point or limit. Thus many of Varela's ideas could now be discussed in terms of the properties of a cartesian closed category, functionals, and elements of objects. Much of the discussion about Goguen's realignment of Varela's calculus has centered on his formation of objects, which can be regarded also as models of the lambda calculus. However I shall focus on an overlooked aspect of their work; that is, the objects, to which Goguen refers, can also be construed as variable sets or objects in a topos of spaces. Indeed we would need to think of them in this fashion, if we desired to theorise using type theory.

Brown's Laws of Form are concerned with the basic act of making a distinction against the backcloth of some indicational space. His Calculus of Indications (CI) has two indicational constants :

- a) \top ; and
- b) \perp , or the empty space, which will occasionally be written as the underscore $_$, for clarity.

Brown's two axioms for these indicational constants are :

$$B1) \top \top = \top ; \text{ and}$$

$$B2) \top \perp = _ .$$

Now Brown, like Smorynski, works in the tradition of those logicians, who believe that "... algebra rarely has anything deep to say about logic" ([SY] p.217). Their basic aim is to study logic in terms of its arithmetic. Thus from these two axioms, we may

calculate the value of any indicational expression (or form) as an arithmetical expression. In arithmetic the sign for equality is usually to be read as the expression on the left has the same value as the expression on the right.

So arithmetically we can calculate :

$$\begin{aligned} \neg\neg\neg &= \neg, \\ \neg\neg &= \neg, \text{ and} \\ \neg\neg &= _ , \text{ by means of (B1) and (B2).} \end{aligned}$$

When variables are introduced, a complete algebra called the Calculus of Indications (CI) can be obtained ([SB]). It has two axioms :

$$\begin{aligned} \text{C1)} \quad \overline{p|q}|r &= \overline{p|r|q|r}| ; \text{ and} \\ \text{C2)} \quad \overline{p|p} &= _ , \end{aligned}$$

where p, q, and r are variables which can be given the value of an indicational constant.

Now CI has an interpretation as the ordinary propositional calculus (PC), in which axiom (C2) has an interpretation as the law of the excluded middle. Table 4 shows how to translate CI into PC.

Interpretation of CI as PC

<u>CI</u>	<u>PC</u>
\neg	true
$_$	false
$p q$	p or q
$\overline{p }$	not p
$\overline{p q}$	p implies q

Table 4

Indeed by translating CI into PC, we can readily prove all of Brown's theorems in the more familiar context of PC. Some have argued, therefore, that formally Brown's calculus adds nothing new. Strait-laced universal algebraists have complained that Brown's notation is ambiguous, in that $\overline{}$ is both a nullary constant and a unary operation ([KP]). I think this criticism is pointless. It is doubtful that any competent mathematical user could be misled by the ambiguity in the notation. Indeed it could be said that the point of the ambiguity is to solve (what could be regarded as) complex theorems in PC by means of a highly visual technique of calculation in CI. The notational ambiguity is a conceptual gain, in that such a well-charted domain as PC could be enriched by considering it as an interpretation of CI.

One of the key innovations in Brown's work is that indicational expressions are relative to one another, since they all stand in relation to some indicational space. Thus in Brown's work an expression like

$$a = \overline{a} \quad (1)$$

is not a valid expression, when interpreted in PC with equality interpreted as logical equivalence. However, (1) may serve as a good representation of a self-referential process, in which the expression 'bites its own tail'. We can capture this circular feedback process by rewriting (1) as

$$a = \overline{} \quad (2),$$

in which the expression (1) re-enters its own indicational space or form. Brown calls such expressions like (2) 'boolean expressions of higher degree'. Strictly speaking, Brown's CI cannot handle such expressions of higher degree, for we have a situation in which either we cannot express a re-entering form or the calculus is

inconsistent. Since Brown's CI has an interpretation as PC, then it is clearly consistent ([KP]). It immediately follows that CI cannot handle the variety of self-referential forms that interested Varela.

Varela saw this failure of CI to meet his objectives as an illustration that self-referential processes could not be represented by non-self-referential ones. He proposed extending the calculus by assigning a specific value to the self-referential domain, by expanding the set of indicational constants to include a sign, $\boxed{}$, standing for self-reference or autonomy. The additional axioms required to incorporate the addition of the new constant were ([VA]) :

$$E1) \quad \boxed{\boxed{}} = \boxed{} \quad ;$$

$$E2) \quad \boxed{} \boxed{} = \boxed{} \quad ;$$

$$E3) \quad \boxed{} \boxed{} = \boxed{} \quad .$$

Similarly, when variables were introduced, an algebra called the Extended Calculus of Indications (ECI) could be obtained ([VA]), in which the following axiom is substituted for (B2) :

$$V2) \quad \boxed{\overline{p} \mid \overline{q} \mid p} = p.$$

Thus the law of the excluded middle is replaced.

Now Varela argued that ECI was functionally complete. Essentially this means that it could be potentially used for 3-valued switching circuits. His claim is false, and it attracted considerable criticism ([KP]). Thus Varela's project of exploring self-reference was in severe conceptual difficulties. Varela acknowledged these difficulties, which appeared to him as two main problems ([VG] p.298) :

- 1) "... the introduction of a third arithmetic value ... will not

really do ... because every re-entering form which is infinite ('viciously' self-referring) takes the same value ... and we can't see the differences as we need to ..."

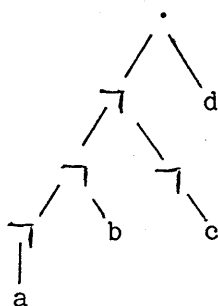
- 2) "... in introducing a third value, certain forms, with a great deal of intuitive meaning, lose this meaning ..."

So Varela concluded that ECI did not represent a satisfactory means of handling self-referential forms. "What ECI provides us with is a map where things went wrong, rather than a map of the forms we wish to express ... re-entry requires a more extensive ground than purely non-re-entering expressions" ([VG] p.298). Goguen was to find that extensive ground in a cartesian closed category.

The key idea, in Varela's joint work with Goguen, is really very simple. Expressions have a very natural ordering vis-a-vis relative degrees of determination and approximation. By dropping the indicational constant, \sqcap , and by using axioms (B1), (B2), (C1), and (V2), a revised ECI is possible. Let F denote the collection of all forms in the revised ECI. Let F_n denote the collection of all forms of depth n or less. Naturally F_n is contained in F_{n+1} . A natural order can now be placed on the collection F of all forms.

Definition 1 Let f, g be any expressions in F . Then g is at least as determined as f , and we write $f \leq g$, if the contents of f coincide exactly with a part (or are equal to) the contents of g when compared, starting at the shallowest depth.

For example, $\overline{a} \mid b \mid \leq \overline{a} \mid b \mid \overline{c} \mid d$. To see this more clearly, it is convenient to display a form as a tree. Thus we display $\overline{a} \mid b \mid \overline{c} \mid d$ as



where '.' denotes the operation of juxtaposition or continece.

So $f \leq g$ when we take both trees and superimpose them (starting at their roots), then the branches of f will coincide exactly with a part of the branches of g . At some points, f will stop, but g will continue to branch further. In this sense, f is less determined than g , or f approximates g by a lesser degree of determination.

Note that the ordering, \leq , in F is not total, for if the roots of two forms are not the same, then they cannot be compared.

Definition 2 Let \perp denote the undetermined expression which approximates everything. Thus $\perp \leq f$, for all f in F .

Proposition 1 The ordering of forms in F is a partial order. That is, $f \leq g$ and $g \leq f$ iff $f = g$.

Proof : ([VG] Prop. 1),

Definition 3 Let f, g be any two forms. The join or least upper bound, denoted by $f \vee g$, is that upper bound of f and g obtained by superimposing f and g at their roots, identifying identical branches and dropping bottoms, where some determined expression occurs at its place. The join is undefined whenever f and g cannot be superimposed.

For example, the join of $\overline{a|b|c}|d$ and $\overline{a|b|c}|d$ is $\overline{a|a|b|c|c}|d$. To be sure, $f \vee g = g$ iff $f \leq g$.

Now consider some sequence of expressions (f_1, f_2, \dots, f_n) , denoted by $\langle f_i \rangle_n$, then the concept of a join can be extended to

that of a supremum over any collection.

Definition 4 The supremum of a sequence $\langle f_i \rangle_n$, when it exists, is a form f such that

$$f = \bigvee_{i=1}^n f_i = \left(\bigvee_{i=1}^{n-1} f_i \right) \vee f_n.$$

This permits consideration of the notion of a limit, by extending joins over countably infinite sequences.

Definition 5 The limit of a sequence $\langle f_i \rangle_\omega$, when it exists, is a form f such that

$$f = \lim_{n \rightarrow \omega} \bigvee_{i=1}^n f_i = \bigvee_{i=1}^{\omega} f_i = \bigvee \langle f_i \rangle_\omega.$$

The collection of all forms, F , is an example of an ω -poset. The only sequences with a limit are those constructed from elements drawn from a countable chain of ascending elements.

Definition 6 A category, \underline{P} , of ω -posets has, for objects, partially ordered sets with a unique bottom (\perp), in which every countable ascending chain of elements has a supremum. The supremum of an ascending chain

$$f_1 \leq f_2 \leq \dots \leq f_n = (f_1, f_2, \dots, f_n) = \langle f_i \rangle_n$$

will be denoted by $\bigvee_n f_n$. The limit of a countably infinite ascending chain will be denoted by $\bigvee \langle f_i \rangle_\omega$. The morphisms of \underline{P} are monotone functions preserving suprema and limits.

Proposition 2 \underline{P} is a cartesian closed category.

Proof : Lambek & Scott ([LS] p.108).

The mathematical reader will readily appreciate that products, order, and suprema are constructed pointwise.

Since F is a poset, it is well-known ([GZ] p.9) that every monotone map

$$\phi : F \longrightarrow F$$

gives rise to a partially ordered set of fixed points

$$\{f \mid \phi(f)=f\}.$$

For every map $\phi : F \rightarrow F$, let

$$\begin{aligned}\phi^0(f) &= f \text{ and} \\ \phi^n &= \phi \circ \phi^{n-1}\end{aligned}$$

for any f in F , then we have the following:

Proposition 3 For any monotone map $\phi : F \rightarrow F$, and a given f in F , $\langle \phi^i(f) \rangle_\omega$ is a chain, with a limit $\bigvee \langle \phi^i \rangle_\omega$.

Proof : ([VG] Prop. 4).

Proposition 4 Every monotone map $\phi : F \rightarrow F$, where F has a unique minimum (\perp), has a least fixed point, $f\phi$.

Proof : ([VG] Theorem 1).

In fact, the least fixed point can be identified with the limit of the following sequence

$$f\phi = \bigvee \langle \phi^i(\perp) \rangle_\omega \quad ([VG] \text{ p.307}).$$

Now since \underline{P} is cartesian closed, and F is a poset with a unique minimum, we can form the following functional

$$\text{fix} : F^F \rightarrow F,$$

which sends a monotone map ϕ to the least fixed point,

$$f\phi = \bigvee \langle \phi^i(\perp) \rangle_\omega.$$

It is now possible to see Goguen's solution to Varela's problems

with ECI. Arithmetical equations, such as $f = \overline{f}$ are to be

identified with a monotone map $\phi : F \rightarrow F$, which sends every

expression f (in the revised ECI) to \overline{f} . A solution to the equation

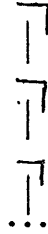
$f = \phi(f)$ is to be found in the least fixed point,

$$\text{fix}(\phi) = f\phi = \bigvee \langle \phi^i(\perp) \rangle_\omega.$$

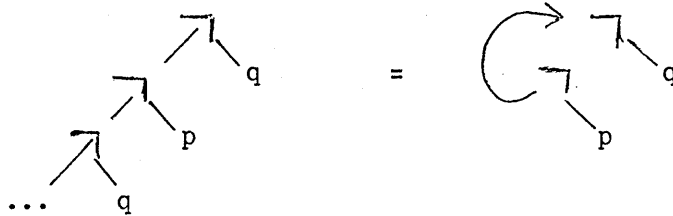
For example, the least fix point solution for $f = \overline{f}$ is

$$\perp \vee \overline{\perp} \vee \overline{\overline{\perp}} \dots = \overline{\overline{\overline{\perp}}},$$

which can be represented as the infinite tree



It is clearly better to represent this as $\overline{\lambda}$ ([VG] p.308). Again consider a monotone map which sends an expression f to $\overline{\overline{f} \mid p \mid q}$. A least fixed point solution to this is ([VG] p.309)



which can be better represented as $\overline{\overline{p} \mid q}$.

In fact, Varela and Goguen show that all re-entrant forms can be viewed as monotone endomorphisms on F , or $F \times F$, or $F \times F \times \dots$ ([VG]).

The cartesian closed category of ω -posets and the functional 'fix' enable Varela to talk about expressions which reduce to λ , $_$, as well as the differentiated re-entrant forms such as $f = \overline{f}$ and $f = \overline{\overline{f} \mid p \mid q}$. The latter are to be identified with the limit of countably infinite sequences

$$(\perp), \phi^1(\perp), \phi^2(\perp), \dots,$$

where ϕ specifies the equation. Varela concludes "... this tells us how it was unnecessary to assume the autonomous state (λ) rather than to construct it" ([VG] p.314). Thus the solution to Varela's problems is one of finding fixed point solutions in a cartesian closed category.

So far I have shown that Goguen has realigned Varela's mathematics with the cartesian closed category \underline{P} . Now \underline{P} is not a

topos, but its objects and morphisms can also be regarded as objects and morphisms in a topos of spaces, namely $\underline{\text{Set}}^{\underline{M}(\omega)^{\text{op}}}$, and it seems to me that this aspect of Goguen and Varela's work can easily be overlooked. They have concentrated on the construction of objects U (in \underline{P}) which are isomorphic to $U \times U$ and U^U . The latter are basically models of the lambda calculus ([LS] p.107), and are useful primarily for relating logic to recursive calculation. There is a danger here of ignoring both the formal aspects of theorising about these objects in terms of type theory and the spatial notions implicit in their work. We could, for example, take a contravariant approach to these objects in \underline{P} , by endowing the posets with the so-called Scott topology ([GZ] p.98). The objects in \underline{P} could now be thought of as topological spaces, and we could investigate their formal properties in terms of sequences, convergence, and limit determined by that topology. However, since we are foresworn to the covariant approach, we may embed \underline{P} in $\underline{\text{Set}}^{\underline{M}(\omega)^{\text{cp}}}$, with topos-theoretic advantage.

Let F be an ω -poset in \underline{P} . The elements of F can be construed as the elements of the set $F(1)$ defined at stage 1 of an object F in $\underline{\text{Set}}^{\underline{M}(\omega)^{\text{op}}}$. Thus an expression f or g (in ECI) is simply a 'point' of $F(1)$. The sequences $\langle f_i \rangle_{\infty}$ drawn from any countably infinite ascending chain in F become the elements of the set $F(I)$ defined at stage I of an object F in the same topos. The internal cohesion of F (considered as a topos object) is such that

- a) for a map $F(t):F(I) \dashrightarrow F(1)$, a sequence $\langle f_i \rangle_{\infty}$ is sent to $\bigvee_t f_t$ for the first t elements in the sequence, and to the limit as t tends to ∞ ;
- b) for $F(m):F(1) \dashrightarrow F(I)$, a 'point' f is sent to the constant

sequence (f, f, f, \dots) ;

c) the monoid action on $F(I)$ picks out the constant sequence equal to the t 'th supremum.

I omit the trivial labour that the above constructions extend to an embedding $\underline{P} \longrightarrow \underline{\text{Set}}^{\underline{M}(\omega)^{cP}}$. Essentially any monotone map $\phi: A \longrightarrow B$ in \underline{P} preserves exactly the same structure that is realised as cohesion in $\underline{\text{Set}}^{\underline{M}(\omega)^{cP}}$.

The justification for taking this covariant approach and treating the collection of forms as an object in a topos of spaces lies in the way that Goguen and Varela want to talk about the revised ECI. They claim ([VG] p.304) that every expression f can be defined as a limit of a sequence, and that for any f they can construct a sequence, $\langle f_i \rangle_n$, that approximates f with any desired degree of accuracy. They report that "... this is quite nice, because there is a neat correspondence between sequences and elements ... this is as much as we need to know about the grounds of completed indicational forms " ([VG] pp.304-5). Thus in order to talk about their problem in any serious way, they need the Kripke-Joyal semantics of $\underline{\text{Set}}^{\underline{M}(\omega)^{cP}}$ to discuss the truth-claims of variable sets of forms at both the stages of forms ($= 1$) and sequences ($= I$). To be sure, they do not consciously in any self-reflective way discuss their problems in terms of topos theory. They remain content to discuss these with respect to the category \underline{P} . However, the 'neat' correspondence between elements and sequences, of which they speak, is no more than the cohesion of an object in the topos $\underline{\text{Set}}^{\underline{M}(\omega)^{cP}}$. The category \underline{P} is not a category of variable sets, and yet their language implicitly discusses objects in \underline{P} as if they were variable sets, as they swap between defining forms in terms of elements defined at the stages of definition of

points and sequences. Thus there seems to be some point in recognising the implicit discussion of variable sets in their work and identifying the topos of spaces which they really need. So I conclude my discussion of the realignment of Varela's notions with a topos of spaces with a further question. What is the smallest subtopos (of spaces) of $\underline{\text{Set}}^{M(\omega)^{op}}$ in which Varela's ECI can live and how does this relate to Johnstone's topological topos ?

CHAPTER TEN

ATKIN AND SIMPLICIAL COMPLEXES

IN SOCIAL STUDIES

Atkin's intriguing attempt to revitalise the use of mathematics in the Social Sciences with concepts borrowed from Algebraic Topology stems from his interest in mathematical physics. He argued that the deployment (in the Social Sciences) of mathematical models borrowed from physics was often superficial and ill-conceived. He thought "... that perhaps Physics is not formulated in such a manner as to show us the heart of its methods, as opposed to the fruits of its labours" ([AT2] p.139). Perhaps a better way forward for the civil application of mathematics in social studies was to study what lay at the 'heart of the methods' of physics, and to transfer to social studies what lay there, rather than to borrow at the surface level of the 'fruits of its (i.e. Physics) labours'.

In some ways, Atkin's search for a methodology that went to the heart of the matter is akin to my own quest for a topos of spaces. He noted that Galileo's profound revolution, in using mathematics as a methodical device, rested on a technical cognitive interest. "Galileo set his mind against the pursuit of 'why' of the motion and concentrated on the 'how'. This required hypotheses ... and experiments to test their validity" ([AT3] p.6). He pointed out that Galileo's experiments required a coordinate framework to reference space-time events ([AT3] pp.6-8). Thus Atkin was led to place great emphasis on geometric figures such as 'points' to

reference position ("...let us use the word 'point' without prejudice" ([AT1] p.200)) and 'paths' to describe motion ([AT3] pp.6-8). Notwithstanding his espousal of Leibnitzian doctrines about relational theories of space ([AT3] p.83), Atkin viewed space-time as a backcloth against which the physicist could make material observations ([AT1] p.191). So "... the property of space-time structure that really matters is its connectivity ..." ([AT1] p.191), which is, perhaps, another way of saying that space-time should have that cohesion which can support motion of matter.

Just exactly what sort of cohesion, Atkin had in mind, can be outlined as follows. The space-time manifolds used by physicists model space (locally) by the mesh diagram of the cartesian coordinates of n -fold copies of the real line R ([AT1] p.200-3). However, Atkin noted that, at best, the measuring instruments of the physicists could only record 'observations' in the rational numbers ([AT1,AT2]). As the mesh diagram is refined, because of the limitations of the measuring technique two unequal points (on the real line) could no longer be distinguished. This lack of distinction defines a neighbour-relation on the set of points referenced by mesh diagram. If point x cannot be distinguished by the measuring technique from point y , then points x and y are neighbour-related, and such a relation is clearly a binary reflexive and symmetric relation. Furthermore, since our techniques of observation have nothing to say about space refined beyond this lack of distinction, we assume the space has the 'geometry' to support motion. Thus if point x is neighbour-related to point y , then it is assumed that there is a 'path' from x to y to support motion. Basically our observations mean that the moving body

travels on boots large enough to span the gaps or holes we cannot observe. Space has the cohesion to support motion. A mathematical physicist, by demanding that the neighbour-relation define an infinitesimal path, could undertake "... the burden of inventing the differential calculus" ([AT1] p.203). The observations with respect to the motion of a body along a path can be found by the process of integrating the changes in observations along the path. Thus, as in equation (7) (Chapter 2), integrating velocity over elapsed time is the distance travelled by the body. Now Atkin argued that these observations by the physicist were such that they portrayed a natural invariant called the cocycle law ([AT1] p.243, [AT2] p.143). Furthermore, there was a relation between the sort of cohesion (or connectivity) that a space has, and the observations which it permits.

We can approach the geometric meaning of the cocycle law as follows. In the first place, a 1-form, ω , is a law (or map) which assigns, to every pair of neighbour-related points x and y , a value in a (multiplicative) group (G, \cdot, e) 'measuring' the 'observation' or 'change' from x to y ([KK3]). For example, if points are related by infinitesimal paths, then $\omega(x, y)$ may be the 'amount of work' needed to move a body the infinitesimal distance from x to y . Of course, we require $\omega(x, x) = e$ (the neutral element of the group G), since no work is required for the null path. If we keep passing from point a to another neighbouring point, then this process traces out a curve or path to a point b . The infinitesimal amounts of work, $\omega(x, y)$, will accumulate (when multiplied) to a value, $\omega(a, b)$, in the group G , which is to be thought of as an integral as follows.

$$\begin{array}{c}
 a \longrightarrow x \longrightarrow y \dots \longrightarrow \dots b \\
 \omega(a,x) \cdot \omega(x,y) \cdot \omega(y,\dots) \cdot \omega(\dots,b)
 \end{array}$$

Under what conditions can we assume the accumulation is an integral ? Recall (Chapter 2, Prop. 1), that every differentiable function on an interval (with values in \mathbb{R}) yields an integral whose value depends only on its value at the endpoints of the curve. Thus in general, for a differentiable function f defined on the interval $[a,b]$ with values in \mathbb{R} , we can set an integral to

$$\omega(a,b) = \int_a^b f' \cdot dt = f(b) - f(a) \quad (1)$$

where f' is the derivative of f .

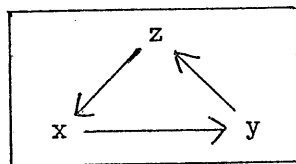
Generalising this to the situation with values in a (multiplicative) group G , we can define a 0-form, γ , as a law (or map) which assigns, to every point x , a value in the group G . We can also define a 1-form, $d\gamma$, which assigns to each neighbour-related pair (x,y) the 'difference' of γ between x and y . Thus

$$d\gamma(x,y) = \gamma(x)^{-1} \cdot \gamma(y) \quad (2).$$

If a 1-form, ω , is equivalent to $d\gamma$ for some 0-form, γ , then ω is said to be an exact 1-form. So one condition for a 1-form to be regarded as an integral might be that it is exact.

In the second place, if the value of an accumulation around an infinitesimal closed curve is the neutral element e , then it is reasonable to expect that the value of an accumulation around a finite null-homotopic closed curve is also e . Our reasons for this expectation are based on the following :

- 1) The value of the amount of work, $\omega(x,x)$ for the null curve must be e , as no work is done.
- 2) The infinitesimal closed curves, which are homotopic to the null curve can be taken to be the triangles (x,y,z) , where the pairs (x,y) , (y,z) , and (z,x) are neighbours.



Since the points x , y , and z are neighbours, it is assumed that the triangle has the geometry to support homotopy to the null curve, since our measuring techniques yield us no knowledge of any 'holes' inside the triangular region bounded by infinitesimal paths. We can define a 2-form, $d\omega$, which is a law (or map), assigning to every triangle (x,y,z) , the value of accumulating the values of ω around the sides of the triangle. Thus

$$d\omega(x,y,z) = \omega(x,y) + \omega(y,z) + \omega(z,x) \quad (3).$$

When $d\omega = e$, then ω is described as a closed 1-form.

Assuming conservation of energy, the value of work done to move an object around an infinitesimal closed curve is to be regarded as accumulating to e , since the work done to return it to its original position cancels out the work done to move it away from that position. So another condition for a 1-form to be regarded as an integral might be that it takes the value e on closed infinitesimal paths.

- 3) The amount of work done around a finite null-homotopic closed curve cancels out in exactly the same way as an infinitesimal closed curve. On the one hand, if null-homotopy is not assumed,

then we may have a closed 1-form which is not exact. On the other hand, we may observe a 1-form which is not closed. In this case, some 'interference' is assumed to alter the amount of work done so that accumulating the amounts of work done around the closed curve no longer cancels to e.

The notion of an integral can be made precise through the cocycle law, which states that all closed 1-forms are exact. That is

$$(d\omega = e) \Rightarrow (d\eta = \omega) \quad (4).$$

Essentially the cocycle law says that space admits integration if closed 1-forms are exact. One of the fundamental theorems of differential geometry says that if space is a connected and simply connected manifold and G is the geometric line ([KK2] p.2), then closed 1-forms are exact ([KK4] p.146). Here connected means that a path exist between each pair of points, and simply connected means that all finite closed curves are homotopic to the null curve. Proofs are provided by Kock ([KK3] p.372).

With the cocycle law, Atkin felt he had arrived at the heart of the methods of mathematical physics. Physical observations were natural invariants, in the sense that observations of 'changes' of a body against the backcloth of space were values in a group such that the pair (backcloth, group of observations) admitted integration. That is, closed 1-forms were exact. But this in turn meant that the backcloths used by the physicist were connected and simply connected (or acyclic). For Atkin, the cocycle law and the acyclic backcloth "... summarise all we can expect from this sort of backcloth. They are amply illustrated in theoretical physics,

albeit somewhat hidden under the various formulations of field theories ..." ([AT1] p.203).

Now it seems to me that Atkin had nearly all the material to hand for developing a topos-theoretic formulation of differential geometry. Working within a suitable topos (in which the axioms of differential geometry were valid), models of the backcloth which admit integration could be exhibited. If it were felt that the cocycle law was of interest to a social scientist as a model of 'sociological observations', then interpretation of that cocycle law in an appropriate topos (lacking the special features of differential geometry) relevant to the social scientist could yield models which may account for such observations. However, this interesting suggestion is only possible with the benefit of hindsight. Atkin formulated his ideas about the relationship between physical observations and the cohesion of the backcloth long before topos-theoretic approaches were widespread amongst pure mathematicians. He attempted a programme of transferring these notions from Physics to Social Science through concepts drawn from algebraic topology. Since I claim that, whenever ideas of space are involved in scientific questions, what we should be reasoning about are objects in some topos of spaces, then it is worth investigating Atkin's programme as a possible alternative to my own.

The most important concept Atkin required from algebraic topology was that of a simplicial complex.

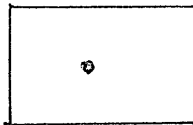
Definition 1 By a simplicial complex, $K(X)$, with vertex set X , we mean a family K of non-empty subsets of X , such that if $A \subseteq X$ belongs to the family K then every non-empty subset of A also

belongs to K . Such subsets in the family K are called simplices and possess a dimension given by

$$\dim(A) = \text{card}(A) - 1 \quad (5).$$

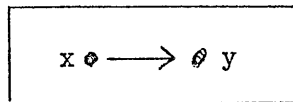
A simplex with dimension p is called a p-simplex. The complex $K(X)$ has dimension m , when m is the maximum dimension of all the simplices of $K(X)$. A subcomplex of $K(X)$ is a family L of simplices of $K(X)$, such that the family L is also a complex $L(A)$ in some subset of vertices $A \subseteq X$.

To be sure, families of sets satisfying the above definition may occur without any reference to geometry. However in his discussions of physics, Atkin had in mind the modelling of spaces with these simplicial complexes. The set, P , of points

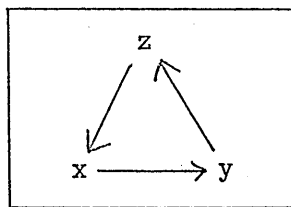


in the space were to be regarded as a vertex set P of 0-simplices.

Neighbour-related points x and y



were to be regarded as 1-simplices, and represented by the 1-simplex $\{x,y\}$, by taking "... away the underlying Euclidean support ..." ([AT1] p.201). Infinitesimal paths composed of neighbour-related pairs (x,y) , (y,z) , and (z,x)



were to be represented by the 1-simplices $\{x,y\}$, $\{y,z\}$, and $\{x,z\}$, and regarded as faces of the 2-simplex $\{x,y,z\}$. A surface would be regarded as a collection of 2-simplices obtained from glueing the triangles together along appropriate edges and points. This process

may be generalised to higher dimensions, but we will not need to deal with this here.

As well as these geometrical models, Atkin was much taken with Dowker's discovery that any binary relation between two sets A and B could be modelled by a pair of simplicial complexes ([DW]).

Definition 2 If $R \subseteq A \times B$ is a binary relation, then $s = \{a_0, a_1, \dots, a_p\}$ is a p -simplex of $K_B(A)$ iff there exists $b \in B$, such that a_i is R -related to b for $i = 0, 1, \dots, p$. Similarly $t = \{b_0, b_1, \dots, b_q\}$ is a q -simplex of $K_A(B)$ iff there exists an $a \in A$, such that a is R -related to b_j for $j = 0, 1, \dots, q$. The complexes $K_B(A)$ and $K_A(B)$ are said to be a pair of conjugate Dowker complexes.

There is a category, SComplex, whose objects are simplicial complexes, and where the morphisms are the simplicial maps $f:K(X) \rightarrow L(Y)$, consisting of functions $f:X \rightarrow Y$, such that if s is a simplex of $K(X)$ then $f(s)$ is a simplex of $L(Y)$. It can readily be seen that the category, SComplex, reduces to one of a simple type of posets and order-preserving maps.

Armed with a technique which could produce a plenitude of simplicial complexes, and an attempted synergy between differential geometry and algebraic topology, Atkin proceeded to argue that the (mathematical) methodical centre of Physics could be transferred to Social Science in the following steps ([AT2] p.159, [AT3] pp.88-90).

Step 1

Set up an observing system and define a binary relation, R , between observed 'phenomena', A , and the 'points', P , of a backcloth.

Working in the category of constant sets, Set, use the relation R to form a Dowker complex $K_A(P)$. In the case of social studies, any 'observer' could observe such a binary relation almost anywhere, and Atkin gives a good selection of these in Mathematical Structure in Human Affairs ([AT3]). In the case of mathematical physics, the selection of a relation is a somewhat trickier affair. Atkin's definition of the complex $K_A(P)$ is somewhat obscure ([AT2] pp.154-5, [AT3] pp.85-90). Yet in his earlier account ([AT1]), it is clear that the 1- and 2-simplices were to be formed from a neighbour-relation on the set of points. It is possible that Atkin could treat the collection of infinitesimal triangles homotopic to the null curve as the set, A , of observed phenomena. In that case, the observed binary relation between phenomena, A , and points, P , would have to be the incidence of points to infinitesimal triangles in order to produce an equivalent complex $K_A(P)$ to that obtained from the neighbour-relation.

Step 2

Observe p -forms to be associated with the p -simplices. In the case of Physics, 1-forms were usually observed and these were expected to obey the cocycle law. In the case of Social Science, integer valued p -forms or 'patterns' were to be observed. These patterns would not necessarily conform with the cocycle law, as the complex $K_A(P)$, when thought of in a geometrical way, might possess 'holes' causing failure of exactness in the observed pattern of closed p -forms. Since the observed patterns were no longer observed against an acyclic backcloth (as was presumed to be the case in Physics), Atkin argued that q -connectivity at dimension q was crucial for understanding the cohesion of the backcloth vis-à-vis observed patterns.

Atkin proposed exploring q -connectivity with the following tools. He says that two simplices, s and t , are q -near iff $\dim(s \cap t) \geq q$. More generally, in a complex $K(X)$, a q -tunnel exists from s to t , when there is a sequence of simplices

$$s = a_0, a_1, \dots, a_n = t \quad (6),$$

such that each a_j in the sequence is q -near to the next for $j = 0, 1, \dots, n-1$. Two simplices, s and t , are said to lie in the same q -component iff they can be joined by a q -tunnel from s to t . The number of q -components is denoted Q_q , and Atkin described an algorithm for calculating the Q -vector (Q_0, Q_1, \dots) . A path in $K(X)$ is clearly a 0-tunnel, from which it follows that Q_0 is the number of path components of $K(X)$ ([GI] p.406).

Atkin claimed that the techniques of analysing the q -connectivity of a Dowker complex, $K_A(P)$, when used in conjunction with the study of q -forms, could yield useful material for theorising about social structure in human affairs ([AT2]). In the next decade, Atkin's ideas about Q -analysis were applied by social scientists (for a survey of applications, see [JS3]). However, the mathematics of Q -analysis was subjected to criticism and refinement by mathematicians ([EJ, GE, GI]). In the early eighties, a Social Science Research Council (UK) discussed the theory and practice of Q -analysis, in which the arguments for and against were rehearsed ([MG]). I do not propose to enter this particular debate here. From my perspective, many of these criticisms of Q -analysis stem from difficulties in relating the Formal to the Conceptual, so it is to these problems I now turn.

Not that questions of the relation between the Formal and

the Conceptual have altogether been absent from those commenting on Q-analysis. Consider the following :

- 1) Griffiths and Evans have observed that Atkin's q-connectivity analysis primarily consists in the listing of q-connected components (and possibly q-forms), followed by a scrutiny and interpretation "... in ordinary language, quite exacting to read, because he now adds detail from ... (elsewhere) ..., to which he has been led by the Q-analysis. It has told him were to look ..." ([GE] p.7). The difficulty here is not that we can associate a Dowker complex with any ordinary language statement that we like, but that a variety of interpretations may suggest themselves. Now Atkin's interpretations (in ordinary language) are often very plausible, but without some form of conceptual control a Q-analyst can always make "... assertions which are not unreasonable but which do not follow from the mathematics as it now stands" ([GE] p.8). Now to a certain extent, it is possible to regard Atkin's practice as unorthodox prescientific work in a raw state ([GE] p.9). Nevertheless, disciplines can only prosper when they are "... based upon rational, scrutinisable argument" ([GE] p.10).
- 2) A similar perspective is stressed by Couclelis. She notes that "... the polyhedra of Q-analysis have strong aesthetic appeal for those ... endowed with a geometric imagination ..." ([CU] p.436). Although we can discuss these in an intuitive fashion, she goes on to suggest that this can be no substitute for "... the logical steps to make scientific argument transparent" ([CU] p.437).

What these examples suggest is the difficulty in Atkin's work of relating what can be exactly said to that which can be precisely conceived. In my view, this is the problem of relating the Formal

to the Conceptual. The ordinary language intuitions need to be formalised then checked rigorously to see that the conceptions are right. Now I do not think that we can follow Couclelis along her road of relating the Formal to the Conceptual, with propositional calculus and boolean algebra ([CU]). Atkin's notions require a more sophisticated level of relating type theory to topoi, and given that notions of 'space' saturate his writings then these topoi will need to be topoi of spaces.

Thus the following questions are raised :

- a) what universe of discourse defines Atkin's practice, and is this a topos of spaces ?
- b) if not, then what struggles have taken place to find such a topos ?

Now Atkin always proceeds as if his universe of discourse was the category SComplex. However a number of authors argue that a simplicial complex is not altogether a suitable representation of the problems referred to by Atkin ([EJ,GE,JS1,SD]). I have not the space to enter into this in any detail here. But basically, when Atkin refers to a Dowker complex $K_A(B)$, he was in the habit of naming the elements of A as a p -simplex $\{b_0, \dots, b_p\}$ when $a \in A$ was R -related to b_i for $i = 0, \dots, p$. Now it is possible that a number of elements of A may name the same simplex ([JS1] p.75, [GE] pp.62-3). Furthermore, a simplex may have 'named' faces, but usually its faces will not be 'named'. For this reason, Seidman has suggested that the category of hypergraphs, Hph, is a more appropriate category for Atkin's universe of discourse ([SD]). A hypergraph is a pair (X, F) where X is the vertex set and F is a family of non-empty subsets of X whose union is X . A member of the family F is

called a hypergraph-edge. Given a binary relation $R \subseteq A \times B$, Seidman claimed that we could form a hypergraph (B,A) , in which each $a \in A$ could be regarded as a hypergraph-edge in the vertices $\{b_0, \dots, b_p\}$, whenever a was R -related to b_i for $i = 0, \dots, p$. Thus a hypergraph can have copies of hypergraph-edges for each a which is R -related to the same set of vertices $\{b, b', \dots\}$. Also we need only add in, as hypergraph-edges, those subsets of vertices we require in any analysis. In a similar vein, other authors have defined Atkin's work in terms of simplicial families, but these would appear to be formally similar to hypergraphs ([EJ] p.377).

Unfortunately the category, Hph, is not a topos. It is not even cartesian closed, for almost the same reason as the category of topological spaces; that is, for a general hypergraph B , it does not seem possible to form an adjunction between products $(_ \times B)$ and exponentials $(_ ^B)$. Thus Hph is not a set-like universe of discourse suitable as the target of an interpretation of a formalised fragment of scientific theory. Hph describes the hypergraph-objects (in Set), which can serve as models of the theory of hypergraphs when interpreted in the topos Set. Earl and Johnson quite properly point out that Atkin's objectives are not the same as developing the theory of hypergraphs. Rather his aims were directed to the development of scientific methodology ([EJ] p.378). However this aim requires the interpretation of a scientific theory in a topos.

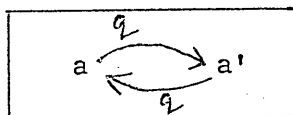
Initial attempts to realign Q-analysis into some topos of spaces stem from theorising in terms of a metaphorical redescription of propulsion of a body along a path as the q-transmission of a q-dimensional object along a q-tunnel ([GE] pp.30-3, [GI] p.407, [JS2]). Consider the motion, m , of a body, B ,

through a space, X , described by

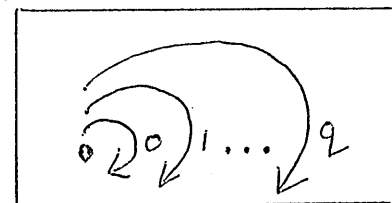
$$B \times T \dashrightarrow X,$$

where T parameterises time. If the dimension of B is q , then q must be less than or equal to the dimension of X . Now imagine the space X has regions with varying dimensions. If U and V are regions of X in which B can be placed, then a q -tunnel is needed to support motion or q -transmission of B from U to V . The regions U and V are said to lie in different q -components if U and V are only connected by a p -tunnel when $p < q$. Thus the existence of multiple q -components is a limitation to communication (or motion / q -transmission) throughout X .

To follow Atkin's movement of thought, Griffiths and Evans ([GE] pp.37-51) have suggested modelling a relation $R \subseteq A \times B$ with a labelled graph (see also Earl & Johnson [EJ]). Call the set A , the nodes of a graph, G . If a and $a' \in A$ name simplices in the Dowker complex $K_A(B)$ such that a is q -near to a' , then form a directed edge from a to a' with the label q .



The resulting graph, G , is an object in the category \underline{RGraph}/Q . Objects are reflexive graphs with a morphism into the labelling graph, Q , which consists of one node and a non-degenerate loop for each label $q = 0, 1, \dots$

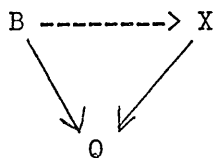


Morphisms of \underline{RGraph}/Q consist of the commutative triangles which

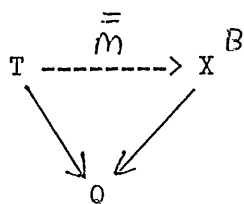
preserve the q -labelling.

In proposing \underline{RGraph}/Q , as a suitable universe of discourse in which to grasp Atkin's ideas, Griffiths and Evans have unerringly elected to work in a gros topos of spaces (although it was not put quite like this). Thus the way is now open to formulate precise scientific theories in type theory and interpret them in this topos. To be sure, this opportunity to use type theory was not taken. However, this realignment of Atkin's mathematics with the simplest sort of space, namely a (labelled) graph, became the primary means for grasping Atkin's ideas. In particular, if X is a q -labelled graph then we can study the subgraphs of X with different labels. In this context, Atkin's q -components identify, as coproducts of q -labelled graphs, those maximal subgraphs where edges are labelled with a value greater than or equal to q . If these subgraphs are not connected then q -transmission cannot be supported between subgraphs. The relation between Q -analysis and graph theory has been fully surveyed by Earl and Johnson ([EJ]).

Nevertheless, the topos \underline{RGraph}/Q may not altogether be a suitable category in which to tease out the problems posed by Atkin's practice. In the first place, mathematical physicists do not need a notion of a q -tunnel. A path or 0 -tunnel is sufficient to support the motion of a particle or 0 -dimensional object of which the body B is formed. Although it is possible to give some substance to the notion of a q -tunnel in physical applications, Griffiths and Evans "... know of no significant applications" ([GE] p.34). Thus working in \underline{RGraph}/Q , the placement of a body B (= graph) can be described by



So we can form the exponential object, X^B , of placements of B in X which preserve the q -label. However difficulties appear when we try to describe motion, m , in terms of the time-dependence of placements,



This would appear to need different copies of the graph representing time, with different labels of q , depending on the motion, m , of the particular body, B . This may or may not be thought 'un-natural'. In the second place, after some empirical investigation, Johnson argued that "... despite a formal development of the theory of q -transmission, it is very difficult to find good examples of q -transmission in real systems ... (since) ... few real systems have q -transmission properties for $q > 0$ " ([JS5] p.294). Given that we have obtained a q -labelled graph from a relation $R \subseteq A \times B$, Johnson argues "... properties which are going to be determined by q alone do not distinguish between vertices (in this case the set B), and this implies some kind of homogenous interpretation of the vertices ... determined by q alone" ([JS4] p.466). He gives a number of reasons why this may prove unacceptable in the social sciences, and these will not be repeated here. Suffice to say that Johnson's analysis moves in the direction of replacing the labelling graph, Q , with a more general graph, V , with values $v \in V$ natural to the application. However a

Q-analysis which is now a V-analysis runs the risk of parting with Atkin's programme. March has argued ([MA]) that many of the features of Atkin's pair of conjugate Dowker complexes can easily be conceived as bicoloured graphs. The import of March's argument is that an appropriate universe of discourse for many of Atkin's ideas is the petit topos of bipartite graphs. Indeed Johnson's later work is set (essentially) in that topos ([JS4, JS5]).

Now it might be argued that, in choosing to work with simplicial complexes, Atkin was already in a gros topos of spaces. A simplicial complex is not just an object of SComplex. It may also be regarded as an object in the category of simplicial sets, $\text{Set}^{\Delta^{op}}$, which is well-known to be a topos of spaces ([KK2] p.293). I do not think this can be seriously countenanced. To be sure, it is open to Atkin to interpret any theories in $\text{Set}^{\Delta^{cp}}$, but there is no evidence to suggest this was seriously considered. Indeed, Griffiths points out that p-forms would have to be specified by more than numerical values. "One would expect a requirement that it have a structure compatible with that of a ... (simplicial complex)" ([GI] p.422). Now I think there are grounds for associating the question of compatible structure with the cohesion of objects in a topos of spaces. So if Atkin were assumed to be working in $\text{Set}^{\Delta^{op}}$, then the group, G, for valuing p-forms would also need to be a simplicial set (= simplicial complex). This is sufficiently absurd to rule out $\text{Set}^{\Delta^{op}}$ as a useful topos for Atkin's practice. A useful way to interpret Griffiths's remark could be to say that simplicial objects and group objects would need to live in the same topos; the compatible structure being that the objects possessed the same kind of cohesion. Thus progress in realigning Atkin's notions with a topos of spaces must issue from

working in such a topos.

Rather than continuing to search for a suitable topos for Atkin's practice, might it not be better to reformulate the cocycle law in a sufficiently general way that it could be interpreted in any topos of spaces useful to a social scientist ? I suggest we recast Atkin's ideas about simplicial complexes in terms of simplicial objects in a topos of spaces. When we do this, it is possible to redefine Atkin's programme along the following lines.

Step 1

Choose an object, X , in some topos of spaces, \underline{E} , which can serve as a backcloth against which our experience of patterns made by an object (of \underline{E}), B , can be observed. Thus we might register our experience of the observed covariation of elements of B with elements of X . For example, physicists might need a topos in which the backcloth possessed geometric figures such as 'points', 'paths', and 'infinitesimals' to support the observed covariation of space-time events. On the other hand, social scientists might require something simpler, such as a graph in RGraph, to support a metaphorical redescription of social structure as a space. Accordingly, Atkin's ambiguous notion of a binary relation between phenomena and geometric points is replaced by the covariant approach to geometric figures in a topos of spaces.

Step 2

Atkin points out that 'nearness' is not just a mathematical concept, but is also deeply embedded in the process of 'observation' ([AT3] p.85). In the case of the differential geometry used by the physicist, a point x is 'near' or neighbour-related to a point y if x and y are not distinct modulo the

observational technique; that is, connected by an infinitesimal path ([AT1], see also similar notions in Beck [BK]). Furthermore, this neighbour-relation, denoted by N , is 'canonical', in the sense that any map $f: X \rightarrow Y$ between manifolds preserves this property of being neighbours ([KK3] p.364) :

$$x N x' \quad \Rightarrow \quad f(x) N f(x') \quad (7).$$

Thus we are led to define a neighbour relation as a binary (not necessarily symmetric) reflexive relation on the (generalised) elements of the objects of \underline{E} , such that the morphisms of \underline{E} preserve the neighbour relation. Except for two trivial extreme cases, objects in an arbitrary topos do not generally carry any 'natural' relation of this kind. More than anything this indicates that we are working in a topos of spaces, such that 'neighbours' can be observed expressing some aspect of the cohesion of a space. For example, in RGraph we can readily 'observe' a neighbour-relation in terms of the adjacencies of a graph. Two vertices a and b are neighbour-related if there is an edge f from a to b . Similarly, two edges f and g are neighbour-related if their source and targets are also neighbour-related. It is clear that a graph morphism must preserve the adjacency structure of a graph, and such morphisms must preserve the above neighbour-relation. Thus we are led to reject Atkin's ambiguous notion of observing a binary relation between phenomena and geometric figures, which seems to have been introduced merely to facilitate the use of Dowker complexes. Instead, we introduce a neighbour-relation on the elements of the backcloth, and this would appear to be in line with Atkin's earlier procedure ([AT1] pp.200-3).

Step 3

We may use the neighbour-relation, N , to form a simplicial object in \underline{E} ([KK2] p.108) :

$$X \begin{array}{c} \xleftarrow{\partial_0} \\ \xleftarrow{\partial_1} \end{array} X(1) \begin{array}{c} \xleftarrow{\partial_0} \\ \xleftarrow{\partial_1} \\ \xleftarrow{\partial_2} \end{array} X(1,1) \quad (8),$$

where the maps, ∂_i , appearing in (8) describe the 'face' operators : 'omit the i th vertex'. X is the object of 0-simplices, $X(1)$ is the object of 1-simplices, and $X(1,1)$ is the object of 2-simplices. The object of 0-simplices is just the backcloth X . Thus a 0-simplex is just a geometric figure in X defined at various stages of definition.

The object of 1-simplices, $X(1)$, is just equivalent to the monic $N \dashrightarrow X \times X$, defined by

$$N \cong X(1) = \{(x, x') \mid x N x'\} \quad (9).$$

Thus, in RGraph, a pair of neighbour-related edges, (f, g) , would be a 1-simplex defined at the stage of edges. Such a pair would force the truth of N at the stage of definition of vertices. However, a 1-simplex in RGraph is an element of a graph $X(1)$. Note that the objects X and $X(1)$ form a graph-object in E.

The object of 2-simplices, $X(1,1)$, is a subobject of $X \times X \times X$ such that

$$X(1,1) = \{(x, y, z) \mid x N y \ \& \ y N z \ \& \ z N x\} \quad (10).$$

To be sure, such an object is defined at all stages of definition, and I leave it to the reader to picture how a 2-simplex might look in his favourite topos. It is possible to form objects of higher dimensional simplices ([KK2] p.108). But these will not be required here.

A degenerate simplex is one in which two vertices are equal, e.g. (x, y, y) is a degenerate 2-simplex and (x, x) is a degenerate 1-simplex. There is an obvious diagonal map $X \dashrightarrow X(1)$ given by $x \dashrightarrow (x, x)$, which may be thought of as a 'degeneracy' operator. This method of forming a simplicial complex (= simplicial object)

from a neighbour-relation contrasts strongly with Atkin's unmotivated use of Dowker complexes defined only at the stage of 'points'.

Step 4

The object is now to define the cocycle law so that it can be interpreted in a topos of spaces. Obtain a (usually abelian) group object, $(G, ., e)$, in \underline{E} , which is to 'measure' the changes in the patterns formed by B 'observed' against the backcloth X . In physics, G is usually the geometric line used to measure 'quantities' ([KK2] pp.2-5). A social scientist, working in RGraph, will need a graph which is also a group object. It is well-known that there are various ways to represent groups as graphs; however, the meaning of such a graph may well be obscure. Let us grant, for the time being, that such a group object has scientific meaning for the social scientist, and proceed as follows.

Definition 3 A k-form on X with values in G is a morphism

$$\omega : X(1, \dots, 1) \longrightarrow G.$$

(k times)

For 1-forms, we require that whenever $\omega(x, y) = a$ and $y \sim x$ then $\omega(y, x) = a^{-1}$. A k -form is called normalised if its value on a degenerate simplex is e . In practice, we will always work with normalised k -forms. It is not difficult to see that for a normalised 1-form, that

$$\omega(x, y) \cdot \omega(y, x) = e \quad (11),$$

for all x, y in X such that $x \sim y$ and $y \sim x$. When $k = 0$, we usually denote a 0-form by $\gamma : X \longrightarrow G$. Since we are working in a topos, we can easily define

- G^X as the object of 0-forms,
- $G^{X(1)}$ as the object of 1-forms, and
- $G^{X(1,1)}$ as the object of 2-forms.

It is now possible to define the following operators.

$$\begin{array}{lll}
 d^0 : G^X & \longrightarrow & G^{X(1)} \quad \text{by} \\
 \gamma & \longmapsto & d\gamma \quad \text{and,} \\
 d^1 : G^{X(1)} & \longrightarrow & G^{X(1,1)} \quad \text{by} \\
 \omega & \longmapsto & d\omega .
 \end{array}$$

The formation of these objects and operators in a topos of spaces permits an interpretation of the cocycle law :

$$\begin{aligned}
 & (\forall \omega \in G^{X(1)}) [(d^1(\omega) = e) \Rightarrow \\
 & \quad (\exists \gamma \in G^X) (d^0(\gamma) = \omega)] \quad (12),
 \end{aligned}$$

where, by abuse, we denote the constant map from the object of 2-simplices with value e as e .

Objects X , $X(1)$, $X(1,1)$, and G which satisfy the cocycle law are said to be a model of the law. Now Atkin argued that physicists used models (in a smooth topos) satisfying the cocycle law. However, he speculated that the k -forms, observed by social scientists, would not necessarily result in models satisfying the law. Of course, the onus would be on the social scientist to theorise about these forms (in type theory). He suggested that the social scientist may well find closed forms which were not exact, and this could be accounted for by the presence of 'holes' in space ([AT3] Chap.5). A further problem (not mentioned by Atkin) could be the failure of the accumulation of 'changes' around a closed path to result in a closed 1-form. Recent work by Kock reveals that the concepts needed to theorise about these problems are pregroupoids (= principal fibre bundles), Ehresmann connections, curvature, and holonomy ([KK4]).

Now it seems to me that these four steps generalise, to any topos of spaces, the notions that Atkin saw at 'the heart of the

methods' of physics. A good generalisation does not search for the maximum generality, but for the right generality. Perhaps the right generalisation for a social scientist, wishing to use mathematics as a methodical device, is that physicists use topoi in which geometric figures cohere. As it is, Atkin's perceptions (when viewed through these four steps) are a particular model in which cohesion is related to 'measure' through a neighbour-relation. We have a supply of models and, as yet, no demand from the Social Sciences. A clear difficulty for a social scientist will be to relate his social scientific concerns to abstract measure theory with a group object (in a topos) for valuing p-forms. What is certain is that we cannot begin to think seriously about Atkin's programme until we realign the mathematics in terms of a topos of spaces. When this is done, a supply of properly formulated models becomes available, and perhaps this may stimulate social scientists to reinvestigate his programme. A critical question must be: 'will these models be suitable for GRA-type studies as well as QC-type studies ?'.

CHAPTER ELEVEN

CONCLUDING COMMENTS

The dictum, that whenever the concept of space enters Science in a fundamental way then our discourse should be interpreted in a topos of spaces, cannot fail to have its practical implications. The claim, that one ought to be in such a topos, is normative and this is a practical matter. As such, the claim is a regulative speech act oriented towards reaching an understanding about norms. The validity claim raised is about the 'rightness' of observing the norm encapsulated in the dictum. To be sure, the 'rightness' of a norm can be countered by force of will. However, my claim is not posed in that sphere of strategic action oriented towards power, but in the realm of the theoretical discourse of Science whose concerns are the rational understanding of the World. If Science is the rational reorganisation of knowledge achieved through a practice that alters its present form then the good reasons for amending that practice need to be judged with respect to the reasonableness to which Science aspires.

A part of those good reasons for changing current practice from a contravariant to a covariant approach rests on Lambek and Scott's adjunction between type theories and topoi. In my view, Formal Logic is a science which systematically attempts to reconstruct and explore the intuitive reasoning powers of competent subjects. Science is couched in constative speech acts which assert the claim to truth. The locutionary element in such speech acts is the content of some proposition, whose truth claim is asserted.

However, formal relations (such as validity, derivability, and consistency) between propositions are the subject matter of Formal Logic, which tries to explicate the mechanisms (and puzzles) of human reasoning. Furthermore, the model theory of Formal Logic attempts to elucidate what can be counted as a model of a (formal) theory. Thus model theory can be regarded as one aspect of a rationally reconstructed knowledge explicating concepts. The adjunction between type theory and topoi tells us that scientific theories with propositional content can be interpreted in topoi. The latter are the subject matter of the Formal and serve as the universe of discourse in which one may find models. Now it might be claimed that my argument places too much weight on this adjunction as the relationship between the Formal and the Conceptual needed for Science. To be sure, there are other well-known relationships between the Formal and the Conceptual. For example, there is an adjunction between cartesian closed categories and typed λ -calculi ([LS]). Although Science needs cartesian closure for elementary calculations (see Chap.8), this is not enough. If theory is to take on the character of universal, existential, and unique quantification typical of the law-like hypotheses of Science then a more powerful version of the Conceptual than typed λ -calculi is needed. The most powerful version of the Formal available to us is type theory, whose appropriate opposite in the Conceptual is Topoi. Now it may be possible that even more powerful versions of the Formal and the Conceptual may be developed which could amend our practice. However, this is speculation. Science is always in the historical situation of depending on the Last Theorist ([HA5] p.277). The truth of tomorrow is an empty concept, and "... we cannot simultaneously assert a proposition or defend a theory and

nevertheless anticipate that its validity-claims will be refuted in the future" ([HA5] p.277). Thus if type theory is the best rational reconstruction of reasoning about the propositions of Science available to us, and Topoi is its opposite in the Conceptual, then there are good reasons for claiming it is 'right' to recast one's mathematics in a topos. There is a rationally grounded norm : interpret theories in a topos.

When 'space' enters the consideration of a theory, then the injunction can be tightened up : interpret the theory in a topos of spaces. Another part of the good reasons for amending current practice is that the rational form of sequential spaces and manifolds used by Science lies in models of the axioms for a *gros* topos. It is not so much that the contravariant approach is 'wrong', but that it obscures the importance of cartesian closure in function space construction and fails to recognise the importance of choosing a topos with the geometric elements that one actually needs. Thus an amended practice stems from the recognition of the importance of exponentiable objects (= function spaces) and geometric figures in indexing objects. The defining feature of 'spaces' are that they are variable sets in which geometric figures cohere (= restrict). They are well-suited to a Science that observes and theorises about the covariation of space-time events. Now it might be argued that a definition of a topos of spaces as 'variable sets in which geometric figures cohere' is insufficiently precise from a mathematical perspective. In Lawvere's view, the difficulties in axiomatising the categories to be counted as 'spaces' stem from "... the lack of a stabilized definition of morphisms appropriate to categories of spaces in the way that 'geometric morphisms' are appropriate to generalized spaces" ([LWG]

p.179). It may well be that clarification of the notion of a 'spatial morphism' could lead to elegant mathematical theories about topoi of spaces. Indeed it would be nice if there were one set of axioms for such topoi. Again the truth of tomorrow is an empty concept and this speculation must be set aside. The differential calculus was used for two centuries before it was 'reformed' by Weierstrass. Similarly there are no grounds for delay in introducing the present (perhaps clumsy) attempts to reform practice through a covariant approach to topoi of spaces. What we do have are various attempts to formalise the idea of a topos of spaces, in which the concept of a gros topos is primary (or, if you prefer, is the earliest stage of definition). There are no a priori grounds for supposing that the idea could be encapsulated in just one set of axioms. The current ideas seem to be just good enough to amend practice, but the current ideas depend on the primacy of the gros topoi.

However, the pivotal role of the gros topoi in 'generating' other examples of topoi of spaces rests on the argument that the gros topoi arise directly out of the needs of a technical cognitive interest to describe and explain 'motion'. In constructing an equaliser (so to speak) between Mathematics and the Philosophy of Science, I have argued that the gros topoi are the domain in which the constitutive questions of a (quasi-)causal Science can be posed. What I have required from Habermas' and Apel's theory of leading knowledge interests is the schematization (appropriate to a technical cognitive interest) of space, time, and substance which can support a causal explanation through 'geometry'. Not only do these different schematizations lead to an understanding of why

mathematics (in the form of topoi of spaces) is the methodical device of the Natural Sciences, but they also lead to an understanding of the differing grounds on which spatial notions can be introduced into the Human Sciences. From my perspective, the theory of leading knowledge interests contributes to a rational reconstruction of the pragmatic dimensions relating knowledge to practice. Thus the good reasons for amending scientific and mathematical practice to a covariant approach using topoi of spaces is grounded in the felicity by which variable sets with coherent geometric figures serve as the methodical tool for describing and explaining the constitutive questions. The path to topoi of spaces for the Human Sciences rests on the second-order methodological objectifications which can arise in the practical cognitive interest. Although these circumscribe the conditions under which we can use categories of spaces, the demand for 'space' is essentially the same.

To be sure, in basing my arguments on these different schematizations I have constructed this equaliser to secure (hopefully) the agreement of those who would not necessarily agree in toto with the theory of leading knowledge interests. It might be argued that the theory is too broad for the conclusions I draw from it, and that I have skated over important problems in the Philosophy of Science. For example, the idea of a technical cognitive interest is floated on the reduction of the logic of explanation to that of prognosis and control. This presupposes observation-controlled feedback of theory amendment, and this in turn presupposes the objectivity of experience. Yet, the presuppositions of the objectivity of experience have hardly entered, except in a tangential way, into my arguments about the

relation between 'geometric figures' and 'experience'. I shall respond to this in two ways. Firstly, in spite of the fact that the theory of leading knowledge interests does not address itself to evidential questions in any detail, I believe Habermas' and Apel's arguments are robust enough to deal with questions about grounding the objectivity of experience ([AP3,HA5,HA6]). I haven't the space to go into this in any detail here. But simply note that in posing my questions within the framework of Metascience I have assumed we have been dealing with corroborated scientific experience. Secondly, in my view these evidential questions are more appropriately located under the heading of Epistemology rather than Metascience. The choice of a particular representation of geometric elements to reflect the objectivity of experience is more properly the subject of Phenomenology in relation to Geometric Logic rather than that of Metascience. With the latter our focus is on how theoretical knowledge (episteme) can best serve practical intelligence (phronesis), and this has been the framework to guide the relevance of questions.

Another part of the good reasons for changing practice lies in the improved steering of the scientific enterprise itself. The injunction to interpret one's theories (with spatial content) into a topos of spaces is designed to improve scientific practice. There is no need for me to argue (counterfactually) that if knowledge of topoi of spaces had been available to Varela, Atkin, and their coworkers then much time and wasted scientific effort could have been saved. However, the case for an increased reflective understanding that one ought to be in a topos of spaces seems to be an overwhelming one, for cases similar to Varela's and Atkin's

problems continue to arise. For example, Dodson and Lok have elucidated the categorical properties of a series of concrete categories, which might be loosely dubbed 'generalised hypergraphs' ([D02,LK]). Now this is a quite proper activity for mathematicians. However, when these categories are presented as offering "... an alternative to continuum mechanics for modelling real phenomena" ([LK]), claims are being made which encroach on the province of my dictum. Now their categories are not topoi (see, for example, [LK] Cor. 3.1.9), nor are they cartesian closed. In fact most of the categories they discuss are not even complete ([LK] Chap.3). Thus, from the vantage point of my dictum their categories do not form a suitable domain for an interpretation of a scientific theory couched in terms of type theory. Now Lok argues that "... it is painfully clear to any physicist that real measurement can only be carried out with finite precision" ([LK] p.3). She postulates that "... this gives rise to the construction of a mathematical model with built-in local uncertainties within which points are indistinguishable" ([LK] p.3), and the concrete categories developed by both Dodson and herself are presented as fulfilling the requirements of just such a mathematical model. However, the import of my own argument is that structure = logic + internal cohesion. While it seems possible to theorise about imprecise measurement in terms of the logic component, it does not seem possible to substitute 'built-in local uncertainty' for the internal cohesion that a space actually has. For example, the twistor theorists use the internal cohesion of a localic topos to parameterise imprecision ([JZ]). They find no need to build-in uncertainty into their spaces. In my view, theories of imprecise measurement are best formulated in a type theory and then interpreted in a topos of spaces with the geometric figures that

one actually needs. Perhaps the mathematical models presented by Dodson and Lok can be viewed as models of a yet unspecified theory in a petit topos $\underline{\text{Set}}_{\underline{E}}(1)$ for some gros topos \underline{E} over $\underline{\text{Set}}$. Thus when mathematicians suggest useful models for Science, my meta-scientific injunction can serve to raise appropriate questions before decisions are taken to invest time and effort on scientific research.

Our ability to embed sequential spaces and manifolds in topoi of spaces might suggest that mathematicians have been using topoi all the time. There is a grain of truth here. But Reyes in recalling (with nostalgia) the benefits of a liberal engineering education argues that the 'synthetic reasoning' (= Kripke-Joyal semantics for variable sets) of engineers and physicists "... does not fit , without violence, into the 'analytic' or set theoretical type of reasoning which has evolved since the end of the last century ..." ([RY] p.69). Thus the injunction to physicists and engineers to use topoi of spaces has certain implications for the modernisation (= renewal) of the mathematical/engineering syllabus. But a start can be made. The first part of Kock's book ([KK2]) is a locus classicus for such synthetic reasoning in differential calculus. Frölicher's and Kriegl's textbook ([FK]) elaborates cartesian closed categories of smooth vector spaces suitable for those taking a covariant approach in the (quasi-)causal Sciences, with scarcely a whisper of their relationship to the gros topoi. Lawvere has also taken a great interest in relating a covariant approach to advanced experimental work in the paradigmatic science of Continuum Mechanics ([LSB]). It seems certain that topoi of spaces, in one way or another, will make their presence felt in the

Natural Sciences; and given that the quasi-causal sciences demand space on a basis derived from that of the Natural Sciences, then the leakage to the Human Sciences will be inevitable.

Much of current mathematical sociology borrows heavily from graph theory in its attempts to express social arrangements and bonds as links in a network ([BH]). It seems to me that mathematical sociology, as a GRA-type Human Science, has much to gain from reflecting about graphs as objects in a topos of spaces. In the first place, graph theory is too often conceived as exploring the combinatorial connectivities of binary relations. To stress the importance that an idea of 'space' is needed for sociology is to point to the really useful part of graph theory (for sociologists) is that consisting of directed multigraphs (= RGraph) and B-partite graphs for some B in RGraph. In the second place, recognition that topos-theoretic structure gives access to exponential objects (a viewpoint not available to traditional graph theory) should tempt mathematical sociologists to apply their theories, whether about graph automorphisms as role-substitutions or the application of predicates to subsets of graph elements, to graphs of this form. Thirdly, the recognition that graphs are objects in a topos of spaces should stimulate exploring other topoi. I have argued that one should choose a topos with the generic elements that one actually needs. Seidman's and Foster's theories about cohesive social subgroups revolve around the fact that these groups can communicate with their members in two steps ([SF2]). Thus it might be worth theorising in $\text{Set}^{\underline{M}(3)^{op}}$ (which has these figures) rather than constraining one's theorising to RGraph $\approx \text{Set}^{\underline{M}(2)^{op}}$. Fourthly, recognition that mathematics enters mathematical sociology as a secondary methodological

objectification supporting some good-reason-assay should end misplaced talk about testing predictions ([BH] p.240) and the analytical power of graph theory ([BH] p.236). It is customary for mathematical sociologists to bewail, from time to time, the sad plight of their subject. However, the metascientific injunction to be in a topos of spaces, when coupled with an understanding of the conditions under which mathematics can enter a GRA-type Science, should serve to encourage in such a Science appropriate, but more limited, expectations.

Finally, nothing in this thesis suggests that a scientist should not be using any other category or topos. If a scientist wants to experiment with other categories, then he should do so. If he wants to continue naively thinking in a set theoretical way then he should do so. But at some point, a scientist will want to move from prescientific meditations to a formal theory and undoubtedly working in a topos of spaces will ease this move. My claim has been a modest one : if you want to talk (formally) about 'space' then find a topos of spaces.

NOTES

(1) Note to Chapter 1

The injunction to work in a topos is clearly normative. My argument that topoi of spaces encode our experience and knowledge of matter moving in space is a rational reconstruction. Such rational reconstructions can play an important role, both positive and negative, in the process of scientific discovery itself. On the positive side, rational reconstructions may 'automatically' provide solutions for many problems (see Chaps. 9 & 10). On the negative side, they may (but not necessarily) be a brake on progress. On the one hand, it forms no part of my argument that a research worker need be conscious or self-reflective about working in a topos of spaces. If I am right then such work can be realigned with such a topos later. Indeed there may well be some good reason for not subjecting the Conceptual to the Formal at an early stage. Thus the discovery that differential forms are really 'glorified functions' derived from 'amazing right adjoints' emerges from conceptual considerations ([LWP] pp.388-90). These 'amazing right adjoints' were quite unknown to categorical logic and only became subject to the Formal after some reflection ([KK2] pp.117-21). The perspective of 'creativity' will always set limits on any norm. On the other hand, by pushing a precise but inadequate rational reconstruction to an unacceptable conclusion, we may expose the source of the inadequacy. Thus if one has some good reason for not heeding my dictum then this rational reconstruction may yet one day find its sublation.

(2) Note to Chapter 5

The gap between this rather abstract definition of (co)limits and the concrete examples in Set (Def.17 ff.) may seem too wide for

some readers to cross. To bridge this gap I take the case of equalisers. Firstly, I show that Definition 17 gives us precise ideas of what conditions an equaliser should meet. Secondly, I show that an equaliser in Set is defined as in example (c) following Definition 17 and meets these conditions.

(i) Assuming the right adjoint $\lim_{\leftarrow \underline{J}}$ to Δ exists, then for each object B in $\underline{A}^{\underline{J}}$, there is an object $\lim_{\leftarrow \underline{J}}(B)$ (denoted by E say) in \underline{A} , and a natural transformation $\xi(B): \Delta \cdot \lim_{\leftarrow \underline{J}}(B) \longrightarrow B$ in $\underline{A}^{\underline{J}}$, such that (Def.15, Prop.1) for every morphism $\bar{f}: \Delta(A) \longrightarrow B$ in $\underline{A}^{\underline{J}}$, there is a unique morphism $f: A \longrightarrow \lim_{\leftarrow \underline{J}}(B)$ satisfying $\bar{f} = \xi(B) \cdot \Delta(f)$. In the case of equalisers \underline{J} is the category represented by $\circ \rightrightarrows \circ$, and a functor $B: \underline{J} \longrightarrow \underline{A}$ is an object B in $\underline{A}^{\underline{J}}$. B can be represented by a pair of parallel arrows $U \xrightleftharpoons[s]{r} V$ in \underline{A} . A morphism $\bar{f}: \Delta(A) \longrightarrow B$ in $\underline{A}^{\underline{J}}$ can be represented by a pair (f', g) of morphisms in \underline{A} , with $f': A \longrightarrow U$, $g: A \longrightarrow V$, such that $g = r \cdot f'$ and $g = s \cdot f'$. We have

$$\begin{array}{ccc}
 U & \xrightleftharpoons[s]{r} & V \\
 \uparrow f' & \nearrow g & \\
 A & &
 \end{array}
 \quad g = rf' = sf'$$

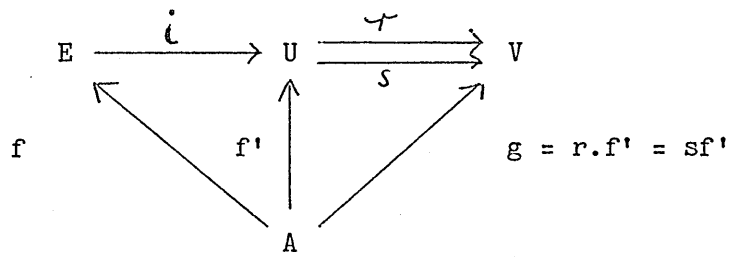
commuting in \underline{A} . Similarly the counit $\xi(B): \Delta \cdot \lim_{\leftarrow \underline{J}}(B) \longrightarrow B$ can be represented by

$$\begin{array}{ccc}
 U & \xrightleftharpoons[s]{r} & V \\
 \uparrow i & \nearrow & \\
 E & &
 \end{array}
 \quad ri = si$$

commuting in \underline{A} .

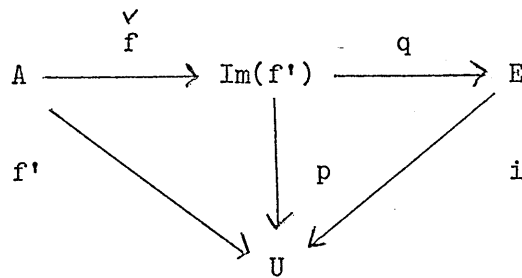
The adjunction means that for any object A in \underline{A} , there is a unique

morphism $f:A \dashrightarrow \varprojlim_{\underline{J}} (B)$ with $\mathcal{E}(B)$. $\Delta(f) = \bar{f}$. This situation can be represented in \underline{A} by



with $f' = i.f$, since $\Delta(f) (= A \xrightleftharpoons[f]{f} E)$ can be represented by $f:A \dashrightarrow E$ in \underline{A} . Thus the task of finding a limit for $B (= U \xrightleftharpoons[s]{r} V)$ reduces to finding a morphism $i:E \dashrightarrow U$ in \underline{A} such that for any $f':A \dashrightarrow U$ (with $sf' = rf'$) then there is a unique morphism $f:A \dashrightarrow E$ with $i.f = f'$.

(ii) In Set, finding an equaliser reduces to finding a set $E = \{x \in U \mid r(x) = s(x)\}$ with $i:E \dashrightarrow U$ the canonical inclusion of E in U . First observe that $ri = si$. Suppose now that $f':A \dashrightarrow U$ is given with $rf' = sf'$, then for every $a \in A$ we have $r(f'(a)) = s(f'(a))$, from which we can see that every $f'(a)$ is an element of E for every $a \in A$, so that the image of f' is contained in E . Let $p:\text{Im}(f') \dashrightarrow U$ and $q:\text{Im}(f') \dashrightarrow E$ be the canonical inclusions. If $\check{f}:A \dashrightarrow \text{Im}(f')$ be described by $a \mapsto \check{f}(a) = f'(a)$ and is the surjection induced by f' then we have



as a commutative diagram. So $i.f = f'$, where $f = q.\check{f}$. Since i is monic, f is the unique morphism with $i.f = f'$, which is what the adjunction requires.

(3) Note to Chapter 6

Schanuel constructed a function $g:D \rightarrow R$ as follows ([KK2] p.5) :

$$\begin{aligned} g(d) &= 0 & \text{if } d &= 0, \\ g(d) &= 1 & \text{if } d &\neq 0. \end{aligned}$$

Note that the function is constructed so that $g.g(d) = g(d).g(d) = g(d)$. Now we cannot be in Set as the Dedekind Reals would mean that $D = \{0\}$ and there would be no unique b to meet the requirements of the axiom. However, by using the law of the excluded middle, we can postulate the existence of $h \in D$, $h \neq 0$, and we have $g(h) = 1$, with $g.g(h) = 1$. But according to Taylor's formula, we must also have

$$g.g(0 + h) = g.g(0) + (g.g)'(0).h.$$

Using Leibniz' Rule, and by elementary calculations ([KK2] p.11) we see that $(g.g)'(0) = g'(0).g(0) + g(0).g'(0) = 0$. So that we have the absurd $(g.g(h) = 1) \neq (g.g(0) + (g.g)'(0).h = 0)$. We can either retain the law of the excluded middle and abandon the axiom, or retain the axiom and abandon the law of the excluded middle. There are good material reasons for retaining the axiom (and its derivations), as scientists use it all the time with spectacular results. So we conclude the differential calculus is a non-classical type theory , in which it is not provable that for all $h \in D$ ($h = 0 \vee h \neq 0$) ([KK2] p.6). In consequence, we must also assume that

$$\{0\} \cup \{h \in D \mid h \neq 0\} \hookrightarrow D$$

is not an isomorphism ([LWP] p.385, Chap.6, Prop.12 ff.). From this perspective, Real Analysis is a different type theory (infinitesimals are abandoned) retaining the law of the excluded middle.

BIBLIOGRAPHY

- AP1 Apel,K-O 1967 Analytical Philosophy of Language and the Geisteswissenschaften (Reidel, Dordrecht)
- AP2 Apel,K-O 1979 "Types of social science in the light of human cognitive interests" in Philosophical Disputes in the Social Sciences (ed) S C Brown (Harvester Press, Brighton) 3-50
- AP3 Apel,K-O 1980 Towards a Transformation of Philosophy (RKP, London)
- AT1 Atkin,R H 1971 "Cohomology of observations" in Quantum Theory and Beyond (ed) E W Bastin (CUP)
- AT2 Atkin,R H 1972 "From cohomology in physics to Q-connectivity in social science" Int. J. Man-machine Studies 4 139-67
- AT3 Atkin,R H 1974 Mathematical Structure in Human Affairs (Heinemann, London)
- BH Barnes,J A & Harary F 1983 "Graph theory in network analysis" Social Networks 5 253-44
- BD Barr,M & Diaconescu,R 1980 "Atomic toposes" J. Pure & Applied Algebra 17 1-24
- BP Barr,M & Pare,R 1980 "Molecular toposes" J. Pure & Applied Algebra 17 127-52
- BW Barr,M & Wells C 1985 Toposes, Triples, and Theories (Springer)
- BK Beck,J M 1977 From the Pocket Calculator to de Rham's Theorem (Aarhus Universitet Matematisk Institut Preprint)
- BN Benabou,J 1985 "Fibered categories and the foundations of mathematics" J. Symbolic Logic 50 10-37
- BS Blass,A 1984 "The interaction between category theory and set theory" Contemporary Mathematics 30 5-29
- BU Bunge,M 1979 "Stack completions and Morita equivalence for categories in a topos" Cahiers de Top. et Geom. Diff. 20 401-35
- CU Couclelis,H 1983 "On some problems in defining sets for Q-analysis" Environment & Planning B 10 423-38
- DH Davis,P J & Hersh,R 1981 The Mathematical Experience (Birkhauser, Boston)
- D01 Dodson,C T 1980 Categories, Bundles, and Space-Time Topology (Shiva, Orpington)
- D02 Dodson,C T 1981 "A new generalization of graph theory" Fuzzy Sets and Systems 6 293-308

- DW Dowker, C H 1952 "Homology groups of relations" Annals of Mathematics 56 84-95
- EJ Earl, C F & Johnson, J H 1981 "Graph theory and Q-analysis" Environment & Planning B 8 367-391
- FA Fay, B 1975 Social Theory and Political Practice (Allen & Unwin, London)
- FI Fine, A 1971 "Relational theories of space" Synthese 22 448-81
- FO Fourman, M P, Mulvey C & Scott D S (eds) 1979 Applications of Sheaves (Springer LNM 753)
- FR Freyd, P 1972 "Aspects of topoi" Bull. Austral. Math. Soc. 7 1-76
- FK Frölicher, F K & Kriegl A 1988 Linear Spaces and Differentiation Theory (Wiley)
- GA1 Gadamer, H-G 1975 Truth and Method (Sheed & Ward, London)
- GA2 Gadamer, H-G 1976 "Hegel and the dialectic of the ancient philosophers" in Hegel's Dialectic: Five Hermeneutical Studies (Yale Univ. Press) 5-34
- GZ Gierz, G et al. 1980 A Compendium of Continuous Lattices (Springer)
- GB Goldblatt, R 1979 Topoi: the Categorical Analysis of Logic (North Holland)
- GJC Gould, P, Johnson J H & Chapman G 1984 The Structure of Television (Pion, London)
- GO Gouldner, A W 1973 "Anti-Minotaur: The myth of a value-free sociology" in For Sociology (Allen Lane, London) 3-26
- GY Gray, J W 1977 "Fragments of the history of sheaf theory" in [FO] 1-79
- GR Greenspan, D 1980 Arithmetic Applied Mathematics (Pergamon Press, Oxford)
- GI Griffiths, H B 1983 "Using mathematics to simplify Q-analysis" Environment & Planning B 10 403-22
- GE Griffiths, H B & Evans P J 1981 Atkin's Q-analysis : a Critique and Reformulation (mimeo Dept. Math. Southampton University)
- HA1 Habermas, J 1972 Knowledge and Human Interests (Heinemann, London)
- HA2 Habermas, J 1973 "Postscript to Knowledge and Human Interests" Philosophy of the Social Sciences 3 157-89
- HA3 Habermas, J 1974 "Introduction: some difficulties in the attempt to link theory and practice" in Theory and Practice (Heinemann, London) 1-40

- HA4 Habermas, J 1976 "The analytic theory of science and dialectics" in The Positivist Dispute in German Sociology (ed) T W Adorno (Heinemann, London) 131-62
- HA5 Habermas, J 1982 "A reply to my critics" in Habermas: Critical Debates (eds) J B Thompson & D Held (MacMillan, London) 219-83
- HA6 Habermas, J 1984 The Theory of Communicative Action I (Heinemann, London)
- HE1 Hesse, M B 1953 "Models in physics" Brit. J. Philosophy of Science 4 198-214
- HE2 Hesse, M B 1963 "The function of models: a dialogue" in [HE4] 7-56
- HE3 Hesse, M B 1964 "The explanatory function of metaphor" in [HE4] 157-77
- HE4 Hesse, M B 1970 Models and Analogies in Science (Notre Dame University Press)
- IS Isbell, J 1986 "General function spaces, products and continuous lattices" Math. Proc. Camb. Phil. Soc. 100 193-205
- JS1 Johnson, J H 1981 "Some structures and notation of Q-analysis" Environment & Planning B 8 73-86
- JS2 Johnson, J H 1982 "Q-transmission in simplicial complexes" Int. J. Man-machine Studies 16 351-77
- JS3 Johnson, J H 1983 "A Survey of Q-analysis" Paper presented to the SSRC Conference on Q-analysis & the Social Sciences at Leeds University 19-20 Sep 1983
- JS4 Johnson, J H 1983 "Q-analysis : a theory of stars" Environment & Planning B 10 457-69
- JS5 Johnson, J H 1986 "Stars, maximal rectangles, lattices: a new perspective on Q-analysis" Int. J. Man-machine Studies 24 293-9
- JT1 Johnstone, P T 1977 Topos Theory (Academic Press, London)
- JT2 Johnstone, P T 1979 "On a topological topos" Proc. London Math. Soc. 38 237-271
- JT3 Johnstone, P T 1985 "How general is a generalized space ? " in Aspects of Topology: Essays in Memory of Hugh Dowker (ed) I M James (London Math. Soc. LN 93) 77-111
- JT4 Johnstone, P T 1987 Notes on Set Theory and Logic (CUP)
- JM Johnstone, P T & Moerdijk, I 1989 "Local maps of toposes" Proc. London Math. Soc. 58 281-305
- JZ Jozsa, R 1979 "Sheaves in physics: twistor theory" in [F0] 492-9
- KI Kitcher, P 1981 "Explanatory unification" Philosophy of Science 48 507-31

- KL Klein, J 1968 Greek Mathematical Thought and the Origin of Algebra (MIT Press, London)
- KK1 Kock, A (ed) 1979 Topos Theoretic Methods in Geometry (Matematisk Institut Various Publications Nu 30, Aarhus Universitet)
- KK2 Kock, A 1981 Synthetic Differential Geometry (London Math. Soc. LN 51, CUP)
- KK3 Kock, A 1982 "Differential forms with values in groups" Bull. Austral. Math. Soc. 25 357-86
- KK4 Kock, A 1986 "Combinatorial notions relating to principal fibre bundles" J. Pure & Applied Algebra 39 141-51
- KR Kock, A & Reyes G E 1979 "Connections in formal differential geometry" in [KK1] 158-95
- KP Kohout, L J & Pinkava, V 1979 "The algebraic structure of the Spencer Brown and Varela calculi" Int. J. General Systems 6 155-171
- LJ Lakoff, G & Johnson, M 1980 Metaphors We Live By (University of Chicago Press)
- LA Lambek, J 1980 "From types to sets" Advances in Mathematics 36 113-64
- LS Lambek, J & Scott, P J 1986 Introduction to Higher Order Categorical Logic (CUP)
- LWS Lawvere, F W 1964 "An elementary theory of the category of sets" Proc. National Academy of Sciences of the U.S.A. 51 1506-10
- LWA Lawvere, F W 1969 "Adjointness in foundations" Dialectica 23 281-296
- LWT Lawvere, F W (ed) 1972 Toposes, Algebraic Geometry, and Logic (Springer LNM 274)
- LWI Lawvere, F W 1972 "Introduction" in [LWT] 1-12
- LWC Lawvere, F W 1976 "Continuously variable sets; algebraic geometry = geometric logic" in Logic Colloquium '73 (eds) H E Rose & J C Shepherdson (North Holland) 135-56
- LWV Lawvere, F W 1976 "Variable quantities and variable structures in topoi" in Algebra, Topology, and Category Theory : a Collection of Papers in Honor of Samuel Eilenberg (eds) A Heller & M Tierney (Academic Press) 101-31
- LWD Lawvere, F W 1979 "Categorical dynamics" in [KK1] 1-28
- LWP Lawvere, F W 1980 "Toward the description in a smooth topos of the dynamically possible motions and deformations of a continuous body" Cahiers de Top. et Geom. Diff. 21 377-92
- LWB Lawvere, F W 1986 "Introduction" in [LSB] 1-16

- LWY Lawvere, F W 1986 "Taking categories seriously" Revista Colombiana de Matematicas 20 147-78
- LWG Lawvere, F W 1986 "Categories of spaces may not be generalized spaces as exemplified by directed graphs" Revista Colombiana de Matematicas 20 179-186
- LWQ Lawvere, F W 1989 "Qualitative distinctions between some toposes of generalized graphs" in Contemporary Mathematics 92 (Amer. Math. Soc. Providence, Rhode Island) 261-300
- LMT Lawvere, F W et al. (eds) 1976 Model Theory and Topoi (Springer LNM 445)
- LSB Lawvere, F W & Schanuel S H (eds) 1986 Categories in Continuum Physics (Springer LNM 1174)
- LK Lok, R W-L 1985 Neighbourhood Systems Ph.D Thesis University of Lancaster
- MG Macgill, S M 1983 "Guest Editorial : Q-analysis : into the second decade?" Environment & Planning B 10 367-70
- ML MacLane, S 1971 Categories for the Working Mathematician (Springer)
- MC McCarthy, T 1978 The Critical Theory of Jürgen Habermas (Hutchinson, London)
- MA March, L 1982 "On Ho's methodological approach to design and planning" Environment & Planning B 9 421-7
- MS Massey, W S 1967 Algebraic Topology: an Introduction (Springer)
- MT Merton, R M 1957 Social Theory and Social Structure (rev. ed. The Free Press of Glencoe, New York)
- MI Mitchell, W 1973 "On topoi as closed categories" J. Pure & Applied Algebra 3 133-9
- MR Moerdijk, I & Reyes, G E 1984 "De Rham's theorem in a smooth topos" Math. Proc. Camb. Phil. Soc. 96 61-71
- NE Nerlich, G 1976 The Shape of Space (CUP)
- OS Osius, G 1975 "A note on the Kripke-Joyal semantics for the internal language of topoi" in [LMT] 349-354
- PA Pare, R 1973 "Connected components and colimits" J. Pure & Applied Algebra 3 21-42
- RA Radnitzky, G 1970 Contemporary Schools of Metascience (2nd Ed, 2 Vols) (Scandinavian University Books, Lund)
- RE Rescher, N 1979 Cognitive Systematization (Blackwell, Oxford)
- RY Reyes, G E 1986 "Synthetic reasoning and variable sets" in [LSB] 69-82

- RI Ricoeur, P 1978 The Rule of Metaphor (RKP, London)
- SE Searle, J R 1969 Speech Acts (CUP)
- SD Seidman, S B 1983 "Rethinking backcloth and traffic: perspectives from social network analysis and Q-analysis" Environment & Planning B 10 439-56
- SF1 Seidman, S B & Foster, B L 1978 "A note on the potential for genuine cross-fertilization between anthropology and mathematics" Social Networks 1 65-72
- SF2 Seidman, S B & Foster B L 1978 "A graph-theoretic generalization of the clique concept" J. of Math. Sociology 6 139-54
- SY Smorynski, C 1985 Self-Reference and Modal Logic (Springer)
- SB Spencer-Brown, G 1969 Laws of Form (Allen & Unwin, London)
- TY Tierney, M 1972 "Sheaf theory and the continuum hypothesis" in [LWT] 13-42
- VA Varela, F J 1979 Principles of Biological Autonomy (North Holland)
- VG Varela, F J & Goguen, J A 1978 "The arithmetic of closure" J. of Cybernetics 8 291-324
- VR de Vries, F-J 1989 Type Theoretical Topics in Topos Theory Ph.D Thesis Rijksuniversiteit te Utrecht
- WA Wall, C T C 1972 A Geometric Introduction to Topology (Addison Wesley, London)
- WE Wellmer, A 1974 Critical Theory of Society (Seabury Press, New York)
- WI1 Wigner, E P 1949 "Invariance in physical theory" Proc. Amer. Phil. Soc. 93 521-6
- WI2 Wigner, E P 1960 "The unreasonable effectiveness of mathematics in the natural sciences" Comm. on Pure & Applied Mathematics 13 1-14
- WR Wraith, G C 1975 "Lectures on Elementary Topoi" in [LMT] 114-206

INDEX OF SYMBOLS

		<u>Page</u>
A, B, C, \dots	categories	72
$A \times B$	product category	76
A/X	slice category over X	76
$A \underline{B}$	functor category	78
$A^{\circ P}$	dual category	78
$A \cong B$	A is equivalent to B	82
$A = B$	A is isomorphic to B	82
A, B, C, \dots	objects	72
	type symbols	97
$A \times B$	product of objects	83
	product type	97
$A \times_C B$	fibred product over C	84
$A + B$	coproduct of objects	83
A^B	exponential object	86
	function type	97
$\text{cod}(f)$	codomain of f	72
$\text{char}(m)$	characteristic of monic m	114
$\underline{C}(A, B)$	class of morphisms from A to B	74
(\underline{C}, J)	site of definition for topology	138
$d\gamma(x, y)$	coboundary of 0-form	193
$d\omega(x, y, z)$	coboundary of 1-form	194
$\text{dom}(f)$	domain of f	72
\underline{E}_j	topos of j-sheaves	138
ev	evaluation morphism	86
$f: A \dashrightarrow B$	morphism	72
	function symbol	97
\overline{f}	unique arrow opposite to f in adjunction	81
\hat{f}	exponential adjoint	86
\overline{f}	name of f	103
		121

f_*	direct image of geometric morphism	135
f^*	inverse image of geometric morphism	135
$f_!$	left adjoint to f^*	153
$f^!$	right adjoint to f_*	153
$f \circ g$ ($f.g, fg$)	composition of morphisms	72
$f \times g$	product of arrows	85
$f + g$	coproduct of arrows	85
$f: \underline{F} \longrightarrow \underline{E}$	geometric morphism	135
(f^*, f_*)	geometric morphism	135
$F: \underline{A} \longrightarrow \underline{B}$	functor	73
h_c, h^c	representable functor	87
$\text{hom}_{\underline{C}}(A, B)$	class of morphisms from A to B	74
Hom	Hom functor	86
$i_j: A \longrightarrow \Sigma A$	injection of jth component	83
$\text{id}_A: A \longrightarrow A$	identity morphism	73
$\text{Id}_{\underline{A}}: \underline{A} \longrightarrow \underline{A}$	identity functor	80
$j: \underline{\Omega} \longrightarrow \underline{\Omega}$	Lawvere-Tierney topology	136
J	Grothendieck topology	137
$\ker(h)$	kernel of character h	114
$K(\underline{C})$	Karoubi envelope of \underline{C}	88
$K(X)$	simplicial complex	196
$K_A(B)$	Dowker complex	198
$\varprojlim \underline{J}$	limit functor	82
$\varinjlim \underline{J}$	colimit functor	82
$L(\underline{E})$	internal language of \underline{E}	126
$L(\underline{\text{Set}})$	internal language of $\underline{\text{Set}}$	103
$\text{Nat}(F, G)$	set of natural transformations from A to B	86
$p_i: A^n \longrightarrow A$	projection onto ith factor	83

$P(A)$	power object	122
	power type	97
$Q q$	q-connected components	200
$R(B)$	reduced category indexed by \underline{B}	170
$S \in P(A)$	variable S of type $P(A)$	98
$\underline{S} \in \underline{E}(B)$	petit topos	170
$\text{true}: 1 \longrightarrow \Omega$	subobject classifier	114
$\text{true}_A: A \longrightarrow \Omega$	$= \text{true} \circ !_A$	114
$T(A)$	topos generated by type theory A	126
$TL(E)$	topos generated by $L(\underline{E})$	127
$x \in A$	variable x of type A	98
Y	Yoneda functor	87
$\gamma: \underline{E} \longrightarrow \underline{\text{Set}}$	'points' functor	135
$\Delta: \underline{A} \longrightarrow \underline{A}^{\underline{I}}$	constancy functor	82
$\Delta: \underline{\text{Set}} \longrightarrow \underline{E}$	'discrete' functor	135
ϵ	counit of adjunction	81
$\eta(x)$	0-form	193
η	unit of adjunction	81
$(\lambda x \in A) \phi(x)$	term of type B^A	102
$\Pi_0: \underline{E} \longrightarrow \underline{\text{Set}}$	'components' functor	153
$\phi(x) \stackrel{x}{=} \top$	validity	105
Ω	truth object	114
	truth type	97
Ω_j	truth object of j -sheaves	138
$\omega(x, y)$	1-form	192
$\nabla: \underline{\text{Set}} \longrightarrow \underline{E}$	'codiscrete' functor	153
0	initial object	83
$0_A: 0 \longrightarrow A$	unique map from initial object	83
$\underline{1}$	discrete category	79
1	terminal object	83
	unity type	97

$!_A : A \longrightarrow 1$	unique map to terminal object	83
$*$	term of unity type	98
$(_)\times B$	product functor	86
$(_)^B$	exponential functor	86
$\langle f_i \rangle_n$	sequence of expressions	183
$\langle f, g \rangle$	product arrow	85
$[f, g]$	coproduct arrow	85
$[\] : H \longrightarrow \underline{E}$	interpretation into \underline{E}	106 132
$[\] : H \longrightarrow L(\underline{E})$	translation into $L(\underline{E})$	132
$\{ \} : X \longrightarrow P(X)$	singleton map	122
\top	true	98
\perp	false or absurdum	98
$--- $	left adjoint relation	80
$ ---X$	entailment relation	99
\models	validity relation	105
\Vdash	forcing relation	129
\neg	logical negation	99
	indicational constant	178
\Box	indicational constant	180
$\bigvee \langle f_i \rangle_\alpha$	limit of sequence of expressions	184